MACHINE LEARNING BASED OPTIMIZATION OF LASER-PLASMA ACCELERATORS

Dissertation

zur Erlangung des Doktorgrades

an der Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Physik

der Universität Hamburg

vorgelegt von

SÖREN JALAS

Hamburg

2023

Gutachter der Dissertation:	Prof. Dr. Wim Leemans
	Dr. Andreas R. Maier
	Prof. Dr. Olle Lundh
Zusammensetzung der Prüfungskommission:	Prof. Dr. Wim Leemans
	Dr. Andreas Maier
	Dr. Jens Osterhoff
	Prof. Dr. Peter Hauschildt
	Prof. Dr. Daniela Pfannkuche
Vorsitzende der Prüfungskommission:	Prof. Dr. Daniela Pfannkuche
Datum der Disputation:	7. Februar 2024
Vorsitzender Fach-Promotionsausschuss Physik:	Prof. Dr. Günter H. W. Sigl
Leiter des Fachbereichs Physik:	Prof. Dr. Wolfgang J. Parak
Dekan der Fakultät MIN:	Prof. Dr.-Ing. Norbert Ritter

Eidesstattliche Versicherung /
Declaration on oath

Hiermit versichere ich an Eides statt, die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt zu haben.

Hamburg, den 16. September 2023,

Sören Jalas

In loving memory of my father.
Your passion for science was a gift that I have cherished. May I honor your memory by igniting that same flame within my own children.

Abstract

Laser-plasma accelerators (LPA) utilize intense laser pulses to drive plasma waves, generating strong electric fields capable of accelerating electrons to relativistic energies. This technology promises accelerator facilities that are orders of magnitude smaller than their conventional counterparts. However, the compactness of LPA systems and the non-linearity of the laser-plasma interaction create a highly coupled environment, where various input and output parameters intricately interconnect. Consequently, manually identifying suitable working points providing the desired beam quality and stability as well as precise tuning of beam parameters to match the demands of specific applications becomes a tedious task, often leading to sub-optimal performance.

The aim of this thesis is to explore the application of machine learning-based methods, particularly Bayesian optimization, within the realm of laser-plasma accelerators. The study involves the implementation of Bayesian optimization to fine-tune the parameters of the LUX accelerator, encompassing simulations and real-time experimentation.

The proposed approach is initially examined through particle-in-cell simulations of the accelerator. After optimizing the system for beam quality, the predictive model built by the optimizer is employed to identify sources of instability and explore potential avenues for improvement. Subsequently, Bayesian optimization is implemented for online control of the experiment. Various methods to address experimental noise and parameter fluctuations are investigated, leading to the identification of techniques emphasizing ultimate beam quality or stability. This way the machine can autonomously tune itself to generate electron beams with sub-percent energy spread and significant charge. Furthermore, a robust operational regime with less than five percent energy spread for 90 % of all shots is identified and studied.

Finally, Bayesian optimization targeting multiple beam parameters enables the identification of tuning curves allowing precise adjustments to specific beam properties while maintaining optimal beam quality. These complex tuning curves describe the intricate balancing of multiple laser and plasma parameters to achieve the most favorable tuning of beam parameters. This approach facilitates charge tuning over nearly 100 pC for various beam energies, while keeping energy spreads below 5 %.

In combination, the methods presented in this thesis provide valuable tools for effectively managing the inherent complexity of LPAs, spanning from the design phase in simulations to real-time operation, potentially paving the way for LPAs to cater to a wide array of applications with diverse demands.

Zusammenfassung

Laser-Plasma-Beschleuniger (LPA) nutzen intensive Laserpulse, um Plasma-Wellen zu erzeugen und starke elektrische Felder zu generieren, die Elektronen auf relativistische Energien beschleunigen. Diese Technologie verspricht Beschleunigeranlagen, die um ein Vielfaches kleiner sind als herkömmliche Linearbeschleuniger. Allerdings führen die Kompaktheit von LPA-Systemen und die Nichtlinearität der Laser-Plasma-Wechselwirkung zu einer stark gekoppelten Umgebung, in der verschiedene Eingabe- und Ausgabeparameter auf komplexe Weise miteinander verbunden sind. Dadurch gestaltet sich die manuelle Suche geeigneter Arbeitspunkte für die gewünschte Strahlqualität und Stabilität sowie die präzise Feinabstimmung auf die Anforderungen spezifischer Anwendungen als mühsame Aufgabe, die oft zu suboptimalen Ergebnissen führt.

Das Ziel dieser Arbeit ist es, die Anwendung von Methoden auf Basis von maschinellem Lernen, insbesondere der Bayes'sche Optimierung, im Bereich der Laser-Plasma-Beschleuniger zu untersuchen. Die Studie umfasst die Implementierung der Bayes'sche Optimierung zur Feinabstimmung der Parameter des LUX-Beschleunigers, sowohl in Simulationen als auch in Echtzeit-Experimenten.

Der verfolgte Ansatz wurde zunächst in Particle-in-Cell-Simulationen des Beschleunigers untersucht. Ein Vorhersagemodell, das bei der Optimierung erstellt wurde, diente dazu, Instabilitätsquellen zu identifizieren und potenzielle Verbesserungswege zu erkunden. Im Anschluss daran wurde die Bayes'sche Optimierung für die Echtzeitsteuerung des Experiments implementiert. Verschiedene Methoden zur Bewältigung von experimentellen Rauschen und Parameterfluktuationen wurden untersucht und Techniken entwickelt, die geziehlt Strahlqualität oder Stabilität optimieren. Auf diese Weise gelang es dem Beschleuniger, sich autonom einzustellen und Elektronenstrahlen mit einer Energiebreite unter einem Prozent bei signifikanter Ladung zu erzeugen. Darüber hinaus wurde ein Betriebsregime identifiziert und untersucht, das eine deutlich verbesserte Stabilität aufwies und für 90 % aller Schüsse eine Energiebreite von weniger als fünf Prozent erzielte. Schließlich wurde die Bayes'sche Optimierung mehrerer Strahlparameter verwendet, um Regelungskurven zu identifizieren, die präzise Änderungen spezifischer Strahleigenschaften ermöglichen und gleichzeitig eine optimale Strahlqualität beibehalten. Diese Kurven beschreiben komplexe Variationen der Parameter des Lasers und des Plasmas, die es ermöglichen, die Ladung über nahezu 100 Picocoulomb für verschiedene Strahlenergien zu variieren, während die Energiebreite unter 5 Prozent bleibt.

In Kombination bieten die in dieser Arbeit präsentierten Methoden wertvolle Werkzeuge, um die inhärente Komplexität von LPAs effektiv zu bewältigen—von der Entwurfsphase in Simulationen bis hin zum Echtzeitbetrieb. Dadurch eröffnet sich das Potenzial mit LPAs, eine breite Palette von Anwendungen mit vielfältigen Anforderungen zu bedienen.

Contents

Publications

The results discussed in this theses have been published in the following articles:

- **S. Jalas**, M. Kirchen, P. Messner, P. Winkler, L. Hübner, J. Dirkwinkel, M. Schnepp, R. Lehe, and A. R. Maier, *Bayesian Optimization of a Laser-Plasma Accelerator*, Phys. Rev. Lett. 126, 104801 (2021) [1]
- **S. Jalas**, M. Kirchen, C. Braun, T. Eichner, J. B. Gonzales, L. Hübner, T. Hülsenbusch, P. Messner, G. Palmer, M. Schnepp, C. Werle, P. Winkler, W. P. Leemans, and A. R. Maier, *Tuning Curves for a Laser-Plasma Accelerator*, Phys. Rev. Accel. Beams 26, 071302 (2023) [2]
- **S. Jalas** and A.R. Maier , *Plasmasurfen auf Pilotanlagenkurs*, Phys. Unserer Zeit, 53: 168-174. (2022) [3]

The following publications include co-authored contributions related to the content of this thesis:

- A. R. Maier, N. M. Delbos, T. Eichner, L. Hübner, **S. Jalas**, L. Jeppe, S. W. Jolly, M. Kirchen, V. Leroux, P. Messner, M. Schnepp, M. Trunk, P. A. Walker, C. Werle, and P. Winkler, *Decoding Sources of Energy Variability in a Laser-Plasma Accelerator*, Phys. Rev. X 10, 031039 (2020) [4]
- M. Kirchen, **S. Jalas**, P. Messner, P. Winkler, T. Eichner, L. Hübner, T. Hülsenbusch, L. Jeppe, T. Parikh, M. Schnepp, and A. R. Maier, *Optimal Beam Loading in a Laser-Plasma Accelerator*, Phys. Rev. Lett. 126, 174801 (2021) [5]
- S. A. Antipov, A. Ferran Pousa, I. Agapov, R. Brinkmann, A. R. Maier, **S. Jalas**, L. Jeppe, M. Kirchen, W. P. Leemans, A. Martinez de la Ossa, J. Osterhoff, M. Thévenet, and P. Winkler, *Design of a Prototype Laser-Plasma Injector for an Electron Synchrotron*, Phys. Rev. Accel. Beams 24, 111301 (2021) [6]
- A. Ferran Pousa, I. Agapov, S. A. Antipov, R. W. Assmann, R. Brinkmann, **S. Jalas**, M. Kirchen, W. P. Leemans, A. R. Maier, A. Martinez de la Ossa, J. Osterhoff, and M. Thévenet, *Energy Compression and Stabilization of Laser-Plasma Accelerators*, Phys. Rev. Lett. 129, 094801 (2022) [7]
- A. Ferran Pousa, **S. Jalas**, M. Kirchen, A. Martinez de la Ossa, M. Thévenet, S. Hudson, J. Larson, A. Huebl, J.-L. Vay, and R. Lehe, *Bayesian Optimization of Laser-Plasma Accelerators Assisted by Reduced Physical Models*, Phys. Rev. Accel. Beams 26, 084601 (2023) [8]

Other contributions that are not directly related to the content of this thesis:

- D. C. Haynes, M. Wurzer, A. Schletter, A. Al-Haddad, C. Blaga, C. Bostedt, J. Bozek, H. Bromberger, M. Bucher, A. Camper, S. Carron, R. Coffee, J. T. Costello, L. F. DiMauro, Y. Ding, K. Ferguson, I. Grguraš, W. Helml, M. C. Hoffmann, M. Ilchen, **S. Jalas**, N. M. Kabachnik, A. K. Kazansky, R. Kienberger, A. R. Maier, T. Maxwell, T. Mazza, M. Meyer, H. Park, J. Robinson, C. Roedig, H. Schlarb, R. Singla, F. Tellkamp, P. A. Walker, K. Zhang, G. Doumy, C. Behrens and A. L. Cavalieri, *Clocking Auger Electrons*, Nat. Phys. 17, 512 (2021) [9]
- M. Kirchen, R. Lehe, **S. Jalas**, O. Shapoval, J.-L. Vay, and A. R. Maier, *Scalable Spectral Solver in Galilean Coordinates for Eliminating the Numerical Cherenkov Instability in Particle-in-Cell Simulations of Streaming Plasmas*, Phys. Rev. E 102, 013202 (2020) [10]
- V. Leroux, S. Jolly, M. Schnepp, T. Eichner, **S. Jalas**, M. Kirchen, P. Messner, C. Werle, P. Winkler, and A. Maier, *Wavefront Degradation of a 200 TW Laser from Heat-Induced Deformation of In-Vacuum Compressor Gratings*, Opt. Express 26, 13061-13071 (2018) [11]
- N. Delbos, C. Werle, I. Dornmair, T. Eichner, L. Hübner, **S. Jalas**, S. W. Jolly, M. Kirchen, V. Leroux, P. Messner, M. Schnepp, M. Trunk, P. A. Walker, P. Winkler, A. R. Maier, *LUX - A Laser-Plasma Driven Undulator Beamline*, Nucl. Instr. Meth. Phys. Res. A 909, 318 (2018) [12]

Previous publications and contributions:

- **S. Jalas**, I. Dornmair, R. Lehe, H. Vincenti, J.-L. Vay, M. Kirchen, and A. R. Maier, *Accurate Modeling of Plasma Acceleration with Arbitrary Order Pseudo-Spectral Particle-in-Cell Methods*, Physics of Plasmas 24, 033115 (2017) [13]
- R. Brinkmann, N. Delbos, I. Dornmair, R. Assmann, C. Behrens, K. Floettmann, J. Grebenyuk, M. Gross, **S. Jalas**, M. Kirchen, T. Mehrling, A. Martinez de la Ossa, J. Osterhoff, B. Schmidt, V. Wacker, and A. R. Maier, *Chirp Mitigation of Plasma-Accelerated Beams Using a Modulated Plasma Density*, Phys. Rev. Lett. 118, 214801 (2017) [14]
- M. Kirchen, R. Lehe, B. B. Godfrey, I. Dornmair, **S. Jalas**, K. Peters, J.-L. Vay, and A. R. Maier: *Stable Discrete Representation of Relativistically Drifting Plasmas*, Physics of Plasmas, 23, 100704 (2016) [15]

Introduction

Particle accelerators, specifically electron accelerators, play a crucial role in a broad array of scientific, medical, and industrial applications, thanks to their ability to generate intense, high-energy beams of electrons. One of the most transformative applications of electron accelerators has been in the realm of light sources, where they have driven major advancements across numerous scientific disciplines. Synchrotron radiation sources [16–18] and free-electron lasers [19–24] are now at the forefront of materials science, chemistry, biology, and medicine, enabling researchers to delve deeper into the microcosmic world than ever before.

Synchrotron light sources produce incredibly bright and finely tuned light, from infrared through X-rays, which can be used to probe the structure of matter at an atomic level. This is instrumental in fields as diverse as semiconductor manufacturing [25–27], where it aids in the design of more efficient microchips, to pharmaceutical research [28–30], where it helps in the detailed study of proteins for drug design. Free-electron lasers (FELs) that leverage relativistic electrons to generate ultrafast, intense X-ray pulses, allow scientists to capture rapid chemical reactions [31, 9, 32] or visualize atomic structures in unprecedented detail [33–35]. Moreover, accelerator based light sources are used to solve challenges of highest societal impact. For instance, in environmental studies [36–38], they can aid in understanding and mitigating climate change by analyzing microscopic particles in the atmosphere and materials capable of storing greenhouse gases.

The above examples present only a selection of today's applications of these devices. As the processes relevant to nearly any technological development continue to scale down to the nanomenter level, the need for high-resolution investigative tools is only expected to grow, cementing the role of accelerators in future scientific, medical, and industrial advancements.

While particle accelerators have been revolutionary in propelling advances across a spectrum of fields, their deployment on a larger scale is often hindered by practical constraints. Traditional accelerators, such as synchrotrons and linear accelerators, are complex, sizeable machines that require significant investments for construction, operation, and maintenance. This inherently limits the number of facilities that can be built and the accessibility of these powerful tools for research and industrial communities.

The size of a linear accelerator is fundamentally determined by two factors: the desired beam energy and the achievable accelerating gradient. In modern radio-frequency based accelerators the accelerating gradient is inherently constrained by material breakdown within the cavities, limiting the maximum achievable electric field to approximately $100\,\mathrm{MV\,m^{-1}}$.

A promising strategy to overcome this limitation involves the utilization of plasma as a medium for the accelerating fields. Given its atomic-level breakdown, a plasma can sustain fields that are several orders of magnitude larger than those achievable in conventional accelerators.

The concept of a laser-driven plasma accelerator was initially proposed by Toshiki Tajima and John M. Dawson in 1979 [39].

The principle of laser-plasma accelerators [40] is grounded in the interaction between intense laser pulses and a plasma medium. The high-intensity laser pulse drives a wake in the plasma, analogous to the wake left by a boat moving through water. This wakefield can generate extremely high electric fields, capable of accelerating charged particles to very high energies over very short distances. The field gradients in plasma accelerators can be thousands of times greater than in conventional accelerators, with values of 100 GV m^{-1} easily achievable. This tremendous increase in the strength of the accelerating fields promises a dramatic reduction in the size of accelerator facilities.

Although plasma accelerators were conceptualized over 40 years ago, their realization as experimental systems with promising beam parameters has been a more recent development. The advent of chirped-pulse amplification [41] and the subsequent revolution in high peak power laser systems provided the necessary technological advancements. It was not until 2004 when three different research groups independently reported [42–44] on the successful generation of narrow bandwidth, high-energy beams from laser plasma accelerators, marking a significant milestone in the field.

The field has witnessed remarkable progress since these initial successes. Through numerous experiments, the laser-plasma accelerator community has collectively demonstrated that these novel devices can generate beam parameters nearly on par with those produced by conventional accelerators. Laser-plasma accelerators have been successful in producing beams with durations in the few-femtosecond range [45], peak currents in the kiloampere scale [46, 47], transverse emittances on the scale of mm mrad [48–50], and energy spreads on the level of 1 % and below [5, 51, 52]. Moreover, there has been a steady increase in achievable beam energy, extending into the GeV range [53–55]. This progressive trend recently culminated in a significant achievement—the production of electrons with an energy of 8 GeV [56].

An essential milestone in this rapid progression was the recent demonstration of a free-electron laser driven by a laser-plasma accelerator in the extreme ultraviolet range [57]. This accomplishment not only affirms the maturity and capabilities of laser-plasma accelerators but also stands as concrete evidence that these novel devices can truly deliver on their promise. By successfully generating high-quality electron beams that can drive free-electron lasers, laser-plasma accelerators have effectively begun to carve out their own niche in the landscape of particle accelerators, underscoring their potential to play a pivotal role in advancing scientific exploration and innovation in the years to come.

Despite the impressive strides made in the field of laser-plasma accelerators, there still exist significant hurdles to their widespread adoption and readiness for user applications. Among these challenges are the issues of stability, reproducibility, and controllability. Specifically, these involve identifying a setpoint that yields optimal parameters, ensuring that these parameters can be consistently reproduced over a majority of shots, and effectively fine-tuning the setpoint to match the demands of the intended application. The highly dynamic nature of the laser-plasma interaction and its sensitivity to initial conditions often translate into

fluctuations in the output beam parameters. This uncertainty is further compounded by the fact that the generation and control of the beams require careful orchestration of multiple, interdependent parameters. Consequently, achieving consistent, reliable performance and exercising fine-grained control over the beam properties represent crucial, yet formidable tasks. As such, these factors currently limit the applicability of laser-plasma accelerators for demanding applications where stable, reproducible beam parameters are paramount.

In this context, an emerging solution resides in harnessing the power of advanced computational techniques, namely machine learning. The vast, multidimensional parameter space of laser-plasma interactions invites the use of intelligent, data-driven methods to explore and optimize the operational conditions.

This thesis introduces machine learning methods, specifically Bayesian optimization—a strategy designed to optimize expensive, noise-affected, and potentially non-linear functions—as essential tools in guiding the full life cycle of a laser-plasma accelerator experiment. From the design of the setup using simulations, through the actual real-time operation of the experiment, to the final interpretation of the results using detailed data-driven models, these applied algorithms serve to unlock the full potential of a given setup in terms of beam quality, and illuminate new regimes of the system, revealing, for instance, the stability required for certain applications.

In particular, a comprehensive data-analysis, automated control, and machine learning toolchain was implemented at the LUX accelerator at Deutsches Elektronen-Synchrotron DESY [12, 4] which now enables autonomous Bayesian optimization of the plasma acceleration process on a reliable day-to-day basis. LUX is an experimental setup dedicated to advancing laser-plasma acceleration from the realm of ad hoc lab experiments to that of precision-engineered machines. This aim is followed by integrating modern accelerator technology for beam diagnostics and manipulation, and by implementing a comprehensive control system to support online drift compensation as well as data-driven analysis and decision-making. These features facilitated the first-ever continuous 24-hour operation of a laser-plasma accelerator [4].

Within this context, Bayesian optimization was implemented to autonomously fine-tune the machine and with that enabled the generation of electron beams with sub-percent level energy spreads and application-relevant charges at LUX. Furthermore, it unveiled an operational regime of the machine that provides remarkable stability in the spectral properties of the beam.

In a subsequent step, the use of Bayesian optimization techniques was extended to the broarder scope of accelerator control. Once operational regimes with remarkable beam parameters are established, the next critical step is to ensure the accelerator can adjust these parameters to match the varying requirements of potential applications. To this end, Bayesian methods were deployed to find optimal tuning curves of the LUX accelerator. These curves provide a tool for adjusting critical parameters, such as beam energy or charge, while maintaining the high-quality beam characteristics. Through this approach, the accelerator operation could be finely tuned, offering precise control over the electron beam parameters and preserving

the overall beam quality, thus ensuring the accelerator is responsive and adaptable to a broad range of experimental conditions.

This thesis is organized as follows: Chapter 1 provides an overview of the basic physics underlying laser-plasma accelerators. This is followed by chapter 2, which introduces the concept and methodologies of Bayesian optimization.

Moving from theory to practical implementation, chapter 3 demonstrates how Bayesian optimization can be applied to laser-plasma accelerators through simulation studies.

To demonstrate the real-world applicability of these findings, a series of experiments was conducted, the setup for which is detailed in chapter 4.

The experimental demonstration of Bayesian optimization and the results obtained are presented in chapter 5. The thesis concludes with chapter 6, where Bayesian techniques are employed to derive optimal tuning curves for the beam parameters in a plasma accelerator.

1 Principles of Laser-Plasma Acceleration

This chapter will give an overview of the physics of laser-plasma accelerators. While many readers will already have a good grasp of the key principles, the focus will be on the elements most relevant to the later analyses in this thesis.

The upcoming derivations are not exhaustive in their mathematical detail, but they draw largely from the excellent introductory works by Gibbon [58] and Esarey, Schroeder, and Leemans [40]. The aim is to provide an understanding of the most important aspects of laser-plasma interaction that are necessary for the following discussion in this thesis.

Section 1.1 will focus on the basic principles of laser-plasma interaction. This will be followed by section 1.2, which covers the key concepts of laser-plasma acceleration.

1.1 Laser-plasma interaction

Before diving into the intricacies of laser-plasma accelerators, its worth to take a moment to look at the core elements that define this technology—namely, lasers and plasmas and the interplay between them.

1.1.1 Plasma

The term plasma refers to a state of matter in which a considerable portion of the atoms are ionized, meaning they are separated into their constituent ions and electrons. This unique environment is characterized by high-energy interactions, and distinct behaviors not observed in the more commonly encountered states of matter such as solids, liquids, and gases.

On a microscopic scale, plasmas are composed of charged particles, but when viewed on a larger scale, they appear charge neutral, a characteristic often termed as quasi-neutrality.

Quasi-neutrality arises from the plasma's inherent tendency to shield local field perturbations. Suppose a positive test charge is introduced into a plasma. Over time, the plasma's ions will recede from this charge while electrons will be attracted towards it resulting in a local variation of the charge density. To understand the scale of this effect, consider a hydrogen plasma where the electrons and ions have equal temperature ($T_e = T_i$) [59]. The velocities of the electrons and ions then behave as,

$$\frac{1}{2}m_e v_e^2 = \frac{1}{2}m_i v_i^2 = \frac{3}{2}k_B T_e, \tag{1.1}$$

where v is the average velocity and m the mass of the particles, k_B is the Boltzmann constant.

According to this, the drift velocities of the two particle species are vastly different:

$$\frac{v_i}{v_e} = \left(\frac{m_e}{m_i}\right)^{0.5} \approx 0.02 \tag{1.2}$$

On the timescales considered in the following the ions are practically stationary and are therefore considered immobile.

In a state of thermal equilibrium, the density distribution of electrons, n_e, within a plasma adheres to a Boltzmann distribution [60]. This can be expressed as:

$$n_e = n_i \exp\left(e\phi(r)/k_B T_e\right), \tag{1.3}$$

Here, n_i signifies the density of the immobile ion background, while $\phi(r)$ represents the electrostatic potential generated by the external disturbance.

This combined with Gauss's law, expressed as:

$$\nabla^2\phi = -\frac{\rho}{\epsilon_0} = -\frac{e}{\epsilon_0}\left(n_i - n_e\right) - \rho_{ext}, \tag{1.4}$$

where $\rho_{ext} = Q\delta(r)$ represents the charge density of the test charge, provides the means to derive the following relation:

$$\phi = \frac{Q}{4\pi\epsilon_0 r}\exp\left(-r/\lambda_D\right). \tag{1.5}$$

Equation (1.5) elucidates the plasma's ability to shield the field of the test charge. As we move away from the site of disturbance, quasi-neutrality of the plasma is observed where $\phi \to 0$ as $r \to \infty$.

The characteristic length scale, over which the potential decreases to $\exp(-1)$, is given by:

$$\lambda_D = \left(\frac{\epsilon_0 k_B T_e}{e^2 n_e}\right)^{1/2}, \tag{1.6}$$

This scale, referred to as the *Debye length*, is a critical measure in plasma physics.

The responsiveness of plasma to external perturbations can be quantified by comparing the Debye length to the thermal velocity of the electrons, denoted as $v_{te} = \sqrt{k_b T_e/m_e}$. This comparison allows us to derive the relevant timescale:

$$t_D = \lambda_D/v_{te} = \left(\frac{\epsilon_0 k_B T_e}{e^2 n_e} \times \frac{m}{k_B T_e}\right)^{1/2}. \tag{1.7}$$

An interesting consequence arises when we consider a perturbation introduced by a displacement of an electron slab relative to the ion background, analogous to a capacitor [59]. The plasma would try to neutralize the local potential and therefore such a disturbance would provoke an oscillation of the electron slab around the origin, an oscillation with a distinctive frequency known as the plasma frequency [60]:

$$\omega_p = t_D^{-1} = \left(\frac{e^2 n_e}{\epsilon_0 m_e}\right)^{1/2} \approx 5.6 \times 10^4 \left(\frac{n_e}{\mathrm{cm}^{-3}}\right) \mathrm{s}^{-1}. \tag{1.8}$$

The plasma frequency is a defining factor in the interaction between a plasma and electromagnetic fields, like those produced by lasers. The plasma response depends significantly on the speed of the field's variation.

When the field varies slowly in relation to the plasma's response time, the plasma can adjust to the changes by repositioning its electrons, effectively turning into a mirror that reflects the electromagnetic waves. Conversely, when the field fluctuates faster than the plasma's response time, the electromagnetic waves have the ability to penetrate through the plasma instead of being reflected.

This behavior is dictated by the condition:

$$\omega > \omega_p, \tag{1.9}$$

where ω denotes the frequency of the electromagnetic field. This equation gives rise to a critical plasma density:

$$n_c \left[\mathrm{cm}^{-3}\right] \approx 10^{21} \lambda \left[\mu\mathrm{m}\right]^{-2}, \tag{1.10}$$

Beyond this critical plasma density, electromagnetic waves with a wavelength λ and longer wavelengths are reflected. Plasmas with $n_e < n_c$, which are those of interest for laser-plasma acceleration, are called *underdense*.

1.1.2 Ponderomotive force

Having established a basic definition of a plasma, the second foundational component of a laser-plasma accelerator is the laser pulse. In the previous section we have discussed, that electromagnetic waves with a sufficient frequency can propagate through a plasma. What remains to be discussed is how it interacts with it. For this it is useful to first look a the behavior of a single electron in a light field.

In this case the motion of the electron is governed by the Lorentz force $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$. Expressed by the vector potential of the laser $\mathbf{A}$ so that $\mathbf{E} = -\partial\mathbf{A}/\partial t$ and $\mathbf{B} = \nabla \times \mathbf{A}$ [61], the equation of motion of the electron is:

$$\frac{d\mathbf{p}}{dt} = -e\left[-\partial\mathbf{A}/\partial t + \mathbf{v} \times (\nabla \times \mathbf{A})\right] \tag{1.11}$$

To consider the interaction of light with an electron, we may initially simplify the light field as an infinite train of linearly polarized plane waves. Upon encountering this wave, the electron initiates an oscillation within the light's polarization plane. As long as this oscillation, often termed as *quiver motion*, stays at subrelativistic speeds, the E-field component of the Lorentz force dominates the electron's motion. This results in the electron oscillating back and forth without a net gain in energy or change in position.

However, with a sufficiently strong field, the electron may reach relativistic speeds. At these speeds, the B-field component of the Lorentz force becomes significant in dictating the electron's motion. The transition to relativistic motion can be conveniently described by the *normalized vector potential* $\mathbf{a} = e\mathbf{A}/m_e c$ and its peak amplitude a_0, which is also referred to as the laser strength parameter. Time-integrating eq. (1.11) for the component lying within the field's polarization plane yields:

$$\mathbf{p}_\perp = \mathbf{a} m_e c \tag{1.12}$$

This equation suggests that the quiver motion of the electron becomes relativistic once $a_0 \gtrsim 1$. Thus, the strength of the light field can play a key role in determining the nature of electron motion within the plasma.

The implications of the relativistic quiver motion are such that the electron's motion is no longer restricted to the polarization direction (x). It now includes an additional longitudinal component (z). The motion of the electron can be described by the following relations [58]:

$$p_z \propto \frac{a_0^2}{4}(1 + \cos 2\phi), \tag{1.13}$$

$$p_x \propto a_0 \cos\phi, \tag{1.14}$$

$$z \propto \frac{a_0^2}{4}\left(\phi + \frac{\sin 2\phi}{2}\right), \tag{1.15}$$

$$x \propto a_0 \sin\phi, \tag{1.16}$$

Here, ϕ represents the phase of the light field. In addition to the oscillation in the polarization direction, the electron now exhibits longitudinal oscillation at twice the frequency of the field. Crucially, the electron begins to drift along with the light field, carrying an average momentum that is proportional to $a_0^2/4$.

Of course, the assumption of an infinitely long and wide plane wave does not practically describe a high-intensity laser pulse, such as the ones employed in laser plasma acceleration. These pulses have a finite temporal envelope, indicating that the field gradually intensifies as the laser approaches and diminishes as it recedes.

For an electron subjected to this, it initiates both transversal and longitudinal oscillations and becomes entrained with the laser pulse. As the peak of the laser envelope moves past the electron, the oscillations begin to moderate, and the drift motion also slows down. Ultimately, when the laser pulse has fully passed, the electron comes to rest again. However, it will be longitudinally displaced from its initial position.

Furthermore, actual laser pulses aren't infinitely wide either. A good approximation is that the intensity of a laser pulse in the focal area exhibits a radial fall-off similar to a Gaussian function. When an electron interacts with such a pulse, it initiates its oscillations as before. However, as the electron drifts away from the laser's axis, the field amplitude diminishes. When the field's sign changes, accelerating the electron back towards the axis, the reduced field strength is inadequate to fully restore the electron to its initial position and momentum. Consequently, the electron drifts away from the axis and gets accelerated.

The apparent force generated from the averaged motion within an oscillating field featuring spatial inhomogeneities is often referred to as the *ponderomotive force*. In the non-relativistic regime ($a_0 < 1$), this force can be expressed as [40]

$$\mathbf{F}_p = -m_e c^2 \nabla(\mathbf{a}^2/2). \tag{1.17}$$

This equation implies that the ponderomotive force opposes the spatial gradient of the field's intensity, pushing the electron away from regions of high intensity. Hence, one can analogize the ponderomotive force to a type of light pressure.

1.2 Laser-plasma acceleration

Building on the prior discussion of laser-plasma interactions, this section will detail how these effects can generate macroscopic plasma waves. Subsequently, we will discuss how these plasma waves can be harnessed for the purpose of electron acceleration.

1.2.1 Plasma waves

The previous section discussed how a perturbed plasma seeks to restore charge neutrality, leading to oscillations of the electron density at the characteristic plasma frequency. Laser pulses, through the ponderomotive force, can effectively trigger such disturbances.

When a high-intensity laser pulse (with $a_0 \approx 1$) is focused into an underdense plasma, it pushes electrons away towards the front and the sides through the ponderomotive force, generating a charge separation against the stationary ion background. The resulting electrostatic field

pulls the electrons back towards their origin. However, gaining momentum in this process, the electrons overshoot their initial positions and hence start to oscillate.

This collective response manifests as a train of plasma density waves trailing the laser pulse, oscillating at the plasma frequency. These waves are commonly referred to as the laser wakefield. Given that the wakefield tails the laser pulse, its phase velocity, v_p, matches the group velocity of the laser, $v_g \approx c$. Consequently, this plasma wave exhibits a wavelength of $\lambda_p = 2\pi c/\omega_p$. For densities on the order of 10^{17} to $10^{19}\,\mathrm{cm}^{-3}$ at which plasma accelerators typically operate, the plasma wavelength is between 100 and 10 μm.

The dynamics of the plasma and the resulting electric fields, generated by a laser pulse, can be expressed by one-dimensional cold fluid equations [62, 63]. These equations are best expressed in a frame of reference that co-propagates with the group velocity of the laser v_g. In this frame, the longitudinal and temporal coordinates (ζ, τ) are defined as $\zeta = z - v_g t$ and $\tau = t$ respectively. Additionally, it is assumed that the evolution of the laser occurs on a time scale that is significantly longer than the plasma perturbation. This assumption, known as the *quasi-static approximation*, essentially implies that $\partial/\partial\tau \approx 0$.

The quasi-static approximation and the use of co-moving coordinates allow us to express the plasma fluid quantities solely as functions of the co-moving coordinate ζ. Under these conditions, the plasma's response to excitation by a laser pulse with a normalized vector potential a can be characterized by the following set of equations. These equations define the plasma density n_e, the normalized electrostatic potential $\phi = e\Phi/m_e c^2$, and the longitudinal electric field E_z [58, 40]:

$$n_e/n_0 = \gamma_g^2\beta_g^2\left(\frac{1}{\psi} - \beta_g\right), \quad \frac{\partial^2\phi}{\partial\zeta^2} = k_p^2\gamma_g^2\left(\frac{\beta_g}{\psi} - 1\right), \quad E_z = -E_0\frac{\partial\phi}{\partial\zeta} \tag{1.18}$$

$$\text{with} \quad \psi = \sqrt{1 - \frac{1+a^2}{\gamma_g^2\,(1+\phi)^2}}, \tag{1.19}$$

where $\beta_g = v_g/c$ and $\gamma_p = \left(1-\beta_g^2\right)^{-1/2}$ and $k_p = \omega_p/c$, n_0 is the density of the static ion background, and E_0 the cold wave-breaking limit [64].

To determine the wakefield generated by a laser pulse with an arbitrary envelope, the 1-D fluid equations (1.18) need to be solved numerically. Figure 1.1 presents the results of such a calculation for both a weak ($a_0 = 0.3$) and a strong ($a_0 = 3$) drive laser, each with a Gaussian temporal pulse shape.

In the case of the weak laser, the pulse drives a sinusoidal plasma density wave with a periodicity of λ_p. This in turn generates a sinusoidal potential that consequently gives rise to a longitudinal electric field. This electric field exhibits a phase shift of $\pi/2$ with respect to the plasma density wave. This behavior, observed with weak laser pulses, is commonly classified as the *linear regime*. The linear regime is typically characterized by a normalized vector potential $a_0 \ll 1$.

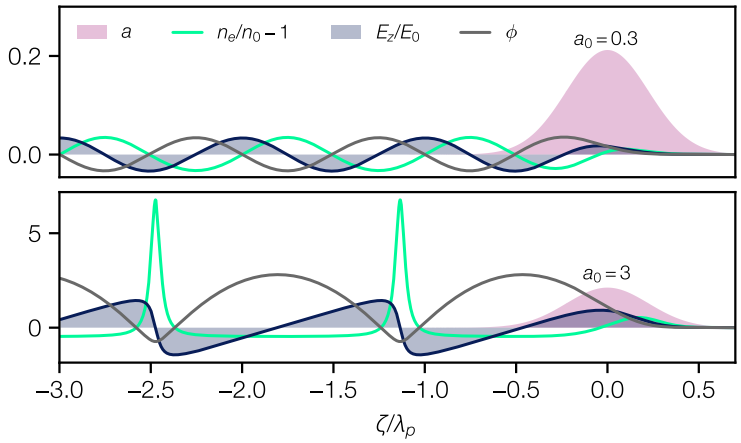

Figure 1.1 – Wakefield excitation of a weak (top, $a_0 = 0.3$) and a strong (bottom, $a_0 = 3$) drive laser. The wakefield is represented in co-moving coordinates $\zeta = z - ct$ that propagate with the drive laser to the right. The fields are found by numerically solving the 1-D cold fluid equations, eq. (1.18). Shown are the envelope of the laser's normalized vector potential a in red, the density response of the plasma $n_e/n_0 - 1$ in green, the longitudinal electric field E_z/E_0 in blue, as well as the normalized electrostatic potential ϕ in grey.

When $a_0 \gtrsim 1$, the quiver motion of the electrons becomes relativistic. This significantly alters the shape of the plasma wave, marking a transition from the linear regime to the *non-linear regime*. As shown in fig. 1.1, the plasma's density response diverges from a sine wave profile. Instead, we observe large regions where the electron density is reduced, followed by a sharp spike in density. This imbalance creates a sawtooth-like electric field, characterized by large areas where the longitudinal field is negative—conditions suitable for electron acceleration.

In the non-linear regime, the wakefield's periodicity no longer aligns precisely with the plasma wavelength. Instead, it extends slightly beyond it. This change is due to the fact that the electrons gain relativistic momentum during their interaction with the laser, leading to an increase in their relativistic mass. As a result of this mass increase, the plasma frequency diminishes causing the wavelength to extend correspondingly. The effective plasma wavelength λ_{np} can analytically be expressed as a function of the peak electric field of the wave $E_{\max}$ [40]:

$$\lambda_{np} = \lambda_p \begin{cases} 1 + 3\left(E_{\max}/E_0\right)^2/16 & \text{for } E_0/E_{\max} \ll 1 \\ (2/\pi)\left(E_{\max}/E_0 + E_0/E_{\max}\right) & \text{for } E_0/E_{\max} \gg 1 \end{cases} \tag{1.20}$$

While an analytical model is available for understanding the three-dimensional linear regime [40], its applicability does not extend to the non-linear case. To capture the complexities of three-dimensional laser-plasma interactions, particularly in the non-linear regime, Particle-in-Cell (PIC) simulations are typically employed [65]. These simulations discretize the plasma into a finite number of macro-particles and the electromagnetic fields are defined on a spatial grid. The dynamics of both the plasma and the fields are modelled by iteratively solving Maxwell's equations for the electromagnetic fields, and the equations of motion for the

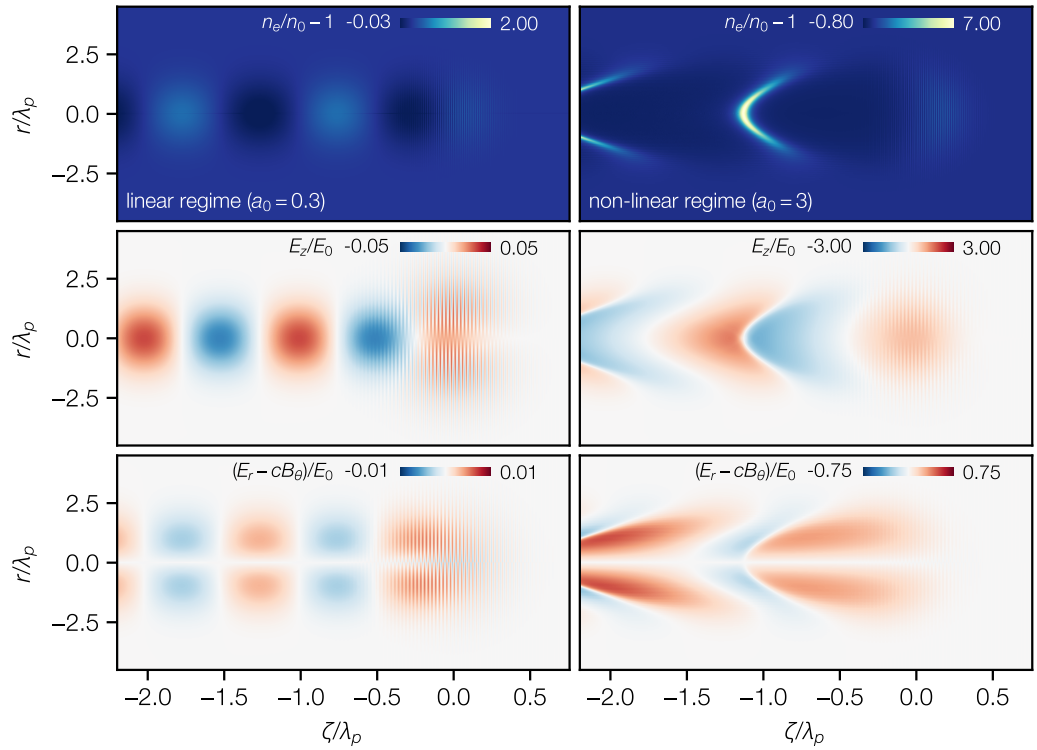

Figure 1.2 – Comparison between linear and non-linear plasma waves. The depicted wakefields are driven by a weak (left) and a strong (right) laser pulse. The unperturbed plasma density n_0 is 10^{18} cm^{-3}, the drive laser has a Gaussian profile, $a = a_0 \exp\left(-\zeta^2/c^2\tau^2\right) \exp\left(-r^2/w_0^2\right)$, with $c\tau = 2/k_p$ and $w_0 = 12/k_p$. Shown are the plasma density $n_e/n_0 - 1$ (top), the longitudinal electric field E_z/E_0 (center), and the focusing fields $(E_r - cB\theta)/E_0$ (bottom). The two cases were modelled with particle-in-cell simulations using the code FBPIC.

particles, which are governed by the Lorentz force. This approach facilitates the simulation of complex systems with minimal assumptions regarding the underlying physics.

Figure 1.2 presents wakefield simulations of the same two scenarios previously investigated using 1-D fluid theory in fig. 1.1. In the linear regime, the sinusoidal behavior observed earlier extends radially, resulting in wavefronts that are orthogonal to the laser propagation axis. As the plasma excitation in the linear regime remains predominantly longitudinal, the radial profile of the wakefield mirrors the profile of the laser pulse. As previously noted, the longitudinal electric field exhibits a phase shift of $\pi/2$ relative to the density modulation. This creates an area behind the drive laser that serves as an acceleration region for electrons (depicted in blue in fig. 1.2). However, this region cannot be entirely filled with electrons due to the focusing fields induced by the plasma wave. Only half of the accelerating region offers fields that focus electrons (illustrated in red in fig. 1.2), while the other half defocus them.

In the non-linear regime, the wakefield shape is significantly different. Consistent with the 1D theory, the regions where the electron density n_e is lower than the background plasma density n_0 elongate, while the density peaks within a very narrow region. In contrast to the linear

regime, these high plasma density fronts are not parallel, but curve towards the drive laser. This curvature is attributed to the relativistic modification of the plasma wavelength. Electrons near the laser's center are exposed to a stronger field, thereby gaining more momentum, and hence result in a longer wavelength than electrons farther from the axis.

The asymmetry of the density modulation in the non-linear regime has a critical consequence: it significantly expands the overlap between regions that accelerate electrons and those that focus them. Specifically, the extensive region of depleted electron density trailing the drive laser—with half of it being suited for electron acceleration (blue in fig. 1.2)—also facilitates stable transport of electrons (red in fig. 1.2).

In instances where the drivers are even stronger than those presented here—typically when $a_0 > k_p w_0$ and $a_0 \gg 1$—the area behind the laser can become completely devoid of electrons, leaving behind only a spherical ion cavity. This scenario is often termed the *blowout* or *bubble* regime. Within this ion cavity, the longitudinal field varies linearly with ζ and remains constant in r. Meanwhile, the focusing fields exhibit linearity with respect to r and retain a constant value in z [66].

1.2.2 Acceleration

Having established an understanding of plasma waves and the regimes of laser wakefields, we now turn our attention to the practical application of these phenomena—the acceleration of electrons using these plasma waves.

Returning to the one-dimensional representation of a non-linear wakefield, we can describe the dynamics of an electron subjected to the longitudinal fields in the wake using the Hamiltonian [67, 68]

$$H(\zeta, p_z) = \left(1 + p_z^2 + a^2(\zeta)\right)^{1/2} - \phi(\zeta) - \beta_p p_z, \tag{1.21}$$

where $p_z = \beta_z \gamma$ is the longitudinal momentum of the electron, γ the corresponding Lorentz factor, ϕ is the wakefield potential, and $\beta_p = v_g/c$ its phase velocity. This Hamiltonian models the electron's total energy, incorporating both potential and kinetic components. Therefore, the principle of energy conservation equates to setting H to a constant value, $H = H_0$. Using this approach, we can derive the evolution of a single electron's momentum, $p_z\left(\zeta, H_0, \zeta_0, p_{z,0}\right)$.

Figure 1.3 gives an impression of the different kind of phase space trajectories (ζ, p_z) an electron can follow depending on its initial conditions. Distinctly, there are two types of trajectories—open and closed. Electrons initially at rest in front of the laser, i.e., ($\zeta_0 \gg 0, p_{z,0} = 0$), follow open trajectories. As they interact with the laser, they acquire some momentum and begin oscillating. They follow the fluid motion that makes up the plasma wave. Over time, these electrons will lag behind, leading to $\zeta \to -\infty$ as $t \to \infty$.

Behind the drive laser, there exist trajectories that form closed orbits. Electrons following these paths are effectively trapped by the wakefield, and they subsequently co-propagate with it. Electrons starting near the back of a wakefield period with low momentum, gradually gain

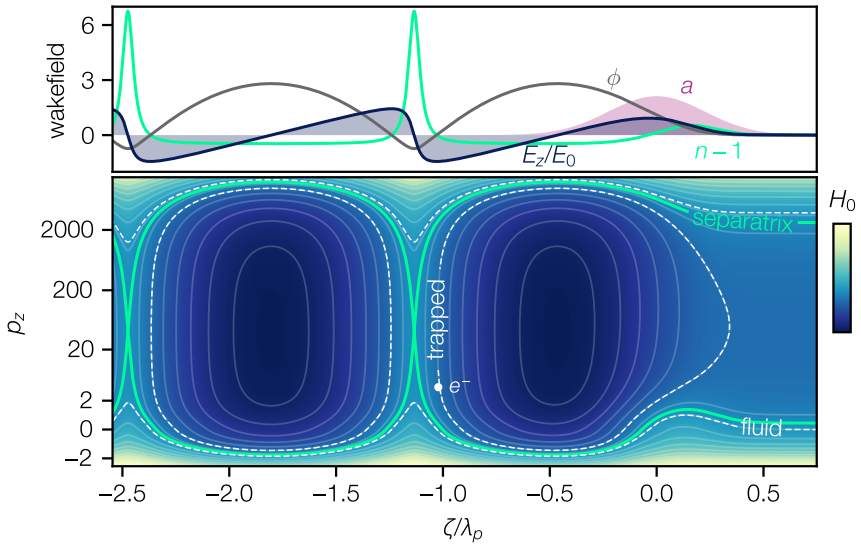

Figure 1.3 – Motion of electrons in a laser wakefield. Top: Wakefield properties according to eq. (1.18) for a driver with $a_0 = 3$. Bottom: Phase space trajectories (ζ, p_z) for different total energy H_0. Figure adapted from [69].

momentum, inching forward towards the drive laser. This movement results from the group velocity of the laser being less than the speed of light within the plasma. However, once the electron reaches the center of the wake period, where the field sign reverses, it begins to lose momentum. Eventually, it lags behind the wake and starts slipping back in phase, triggering another cycle of momentum gain.

The boundary between the closed and open trajectories is denoted by the *separatrix*. This trajectory delineates the maximum possible momentum gain an electron can achieve within a non-evolving wakefield.

1.2.3 Injection

As we have seen, for electrons to take advantage of the immense electric fields of the plasma wave, they must be located on trapped orbits. However, these orbits are not naturally populated, necessitating methods to inject electrons into the wakefield.

One could, in principle, introduce an external electron beam with adequate momentum into the plasma alongside the laser, where it would then be further accelerated. However, this external injection method imposes stringent requirements on the initial electron parameters, essentially necessitating a sophisticated conventional accelerator in addition to the plasma accelerator[1]. Therefore, it is often more desirable to devise methods for direct injection of electrons from within the plasma.

[1] This approach is of cause also relevant when staging multiple laser-plasma accelerators [70].

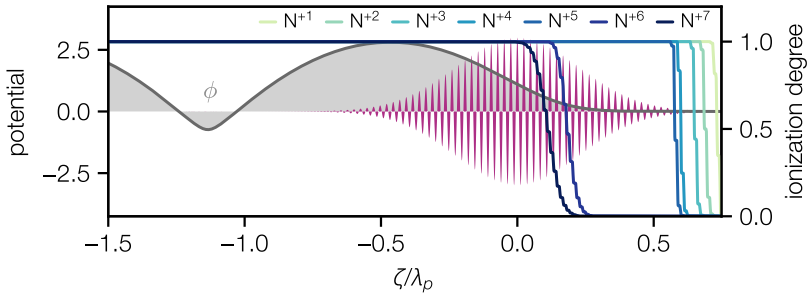

Figure 1.4 – Ionization of nitrogen by a laser pulse with $a_0 = 3$ at 800 nm wavelength. Outer shell electrons are ionized in the leading edge of the laser pulse where the wakefield potential is nearly zero. N^{+6} and N^{+7} start appearing in the core of the laser where the potential is non-zero, the corresponding electrons can be trapped and accelerated. The ionization degree here is modelled according to the ADK-model [77, 78].

Self-injection—When the plasma wave is driven with sufficient intensity, the fluid description of the plasma can break down locally, especially at the high-density spikes at the end of each wake period. Specifically, this happens when the fluid momentum exceeds the phase velocity of the wakefield ($\gamma > \gamma_p$). In such situations, the plasma wave may break in a manner similar to an ocean wave approaching the shore, and electrons may get caught in closed orbits. This self-injection method is the most straightforward way to trap and accelerate electrons. However, the onset of self-injection depends on several factors, including the shape of the wakefield [71, 72], the temperature of the ionized plasma [73], and the precise evolution of the drive laser [74]. Therefore, it can be extremely challenging to predict and control.

Downramp injection—Self-injection can be intentionally initiated by temporarily reducing the phase velocity of the wakefield. One method to accomplish this is by introducing a negative density gradient into the plasma profile [75, 76]. The accompanying elongation of the plasma wavelength as the density decreases causes the wake's wavefronts to lag behind the drive laser. This implies a local reduction of the phase velocity, triggering wave breaking and, subsequently, self-injection.

Ionization injection—While self-injection is based on trapping electrons from the plasma background, the idea behind ionization injection involves the generation of electrons through ionization, in phases of the wakefield where these electrons are already located on trapped orbits. Typically, the plasmas that are used for laser-plasma acceleration are made from gases with a low atomic number like hydrogen. Hydrogen starts to get ionized at an intensity of $1.4 \times 10^{14}\,\mathrm{W\,cm^{-2}}$ [58]. With typical peak intensities on the order of $10^{18}\,\mathrm{W\,cm^{-2}}$, hydrogen atoms are primarily ionized well ahead of the main pulse. As previously discussed, these early-ionized electrons aren't trapped but rather engage in the collective fluid movement of the plasma.

This is no longer the case if a high Z gas like nitrogen is introduced. In this case the inner shell electrons that are more tightly bound are ionized only in the core of the laser pulse.

This scenario is visualized in fig. 1.4. The nitrogen's first five electrons are entirely ionized by the leading flank of the laser pulse. Similar to hydrogen, these electrons cannot be trapped and instead participate in the plasma's collective fluid motion. However, the nitrogen's inner K-shell electrons are ionized solely within the pulse's central region, where the partially formed plasma wave causes the wakefield potential to be non-zero.

In order to be on a trapped orbit, the potential at the phase in which the electron is ionized $\phi(\zeta_{\text{ion}})$ must fullfil the condition [79]:

$$\phi(\zeta_{\text{ion}}) \geq \phi_{\text{min}} - \gamma_p^{-1} + 1, \tag{1.22}$$

where ϕ_{min} is the minimum wakefield potential. In simple terms, this condition means that while the electron starts off slower than the wakefield, it needs to gain enough speed to match the phase velocity of the wakefield by the time it reaches the end of the wake period.

For eq. (1.22) to be fulfilled, the wakefield usually needs to be relatively non-linear with $a_0 \gtrsim 2$ [79]. For a laser driver with a wavelength of 800 nm, the laser field then exceeds the threshold for the $N^{5+} \rightarrow N^{6+}$ ionization process.

The assumption behind eq. (1.22) is that the electron starts from rest. This is plausible if ionization primarily happens at the peaks of the laser field where $a = 0$. However, ionization might occur around these peaks, especially when the laser field surpasses the ionization threshold. The momentum gained in these circumstances is conserved, leading to a degradation of the transverse phase space volume (or emittance) of the electron bunch that's injected [80]. Thus, operation near the injection threshold is often favored. However, control over the injection volume—the region where ionization results in phases of trapped electrons—remains critical for managing both beam charge and quality. Moreover, the trapping condition eq. (1.22) depends on the phase velocity of the wakefield, which means that ionization injection can be combined with downramp injection for precise control over the injected charge.

There are other sophisticated injection methods as well, such as employing multiple colliding lasers [81], or coupling ionization injection with the use of multiple multi-colored lasers [82].

1.2.4 Acceleration limits

With an electron bunch injected the remaining question is how much energy can be extracted from the wakefield. In the ideal case the energy gain of an electron over an acceleration length L_{acc} is

$$\Delta W = eE_z L_{\text{acc}}. \tag{1.23}$$

While the massive fields on the order of several hundred GV m^{-1} might suggest the feasibility of achieving astronomical energy levels with a sufficiently long plasma, practical considerations

impose restrictions on the maximum length of a single acceleration stage. These limiting factors must be taken into account when designing and operating laser-plasma accelerators.

Diffraction—To generate the intensity needed for driving plasma waves, the laser pulse must be tightly focused. However, due to the Rayleigh diffraction, the laser spot size evolves according to $w = w_0(1 + z^2/Z_R^2)^{1/2}$, where w_0 is the size at focus and $Z_R = \pi w_0^2 \lambda^{-1}$ is the Rayleigh length. This means that a more focused laser diffracts faster. For a laser propagating freely, the intensity to generate wakefields is typically maintained only within a few Rayleigh lengths which is typically on the order of millimeters.

Overcoming this limitation involves various methods to guide the laser pulse through the plasma [83–85]. The most relevant approach is to create a plasma channel where the density increases radially. This results in an index of refraction that effectively counteracts the inherent diffraction of the laser.

Dephasing—As we previously discussed in the context of trapped orbits, the electrons gradually advance in phase towards the laser. This is a consequence of the laser's group velocity being lower than the speed of light, while the relativistic electron moves with a velocity $v_z \approx c$. An electron gains energy as long as it's in a region where the electric field is negative, once it reaches the center of the charge depleted region where the field flips—the dephasing point—it starts to lose energy again. For an electron starting at the end of the wake period this happens over the *dephasing length* [86]

$$L_d = \frac{\lambda_p^3}{2\lambda_0^2} \times \begin{cases} 1 & \text{for } a_0^2 \ll 1 \\ (\sqrt{2}/\pi) a_0 & \text{for } a_0^2 \gg 1, \end{cases} \tag{1.24}$$

where λ_0 is the wavelength of the driver. In a dephasing-limited system, the maximum energy gain scales inversely with plasma density $\Delta W \propto n^{-1}$ [40]. However, dephasing effects can be partially offset using tapered density profiles [87]. By gradually reducing the plasma wavelength during propagation, the accelerating wakefield phase can effectively keep pace with the electron.

Depletion—By exciting a wakefield the laser transfers energy to the plasma. The *pump depletion length* is the length after which the laser is fully absorbed by the plasma and can be estimated by equating the total laser energy to the energy contained in the wake behind it $E_z^2 L_{pd} \simeq E_{\text{laser}}^2 L_{\text{laser}}$ [40]. For a square pulse with $L_{\text{laser}} = \lambda_p/2$ the pump depletion length then is [86, 88]

$$L_{pd} = \frac{\lambda_p^3}{\lambda_0^2} \times \begin{cases} 2/a_0^2 & \text{for } a_0^2 \ll 1 \\ (\sqrt{2}/\pi) a_0 & \text{for } a_0^2 \gg 1. \end{cases} \tag{1.25}$$

It's worth noting that for non-linear cases, the pump depletion length (L_{pd}) equals the dephasing length (L_d). However, the pump depletion length is a somewhat drastic metric as it denotes complete laser absorption. As the laser energy depletes, it experiences substantial red-shifting and changes in temporal profile which significantly impacts the shape of the

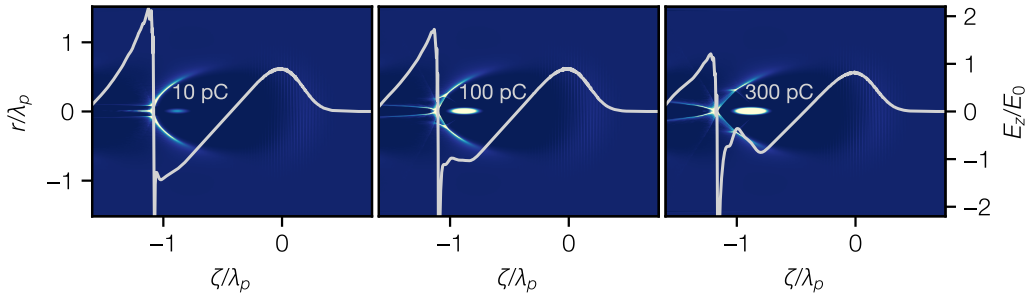

Figure 1.5 – Simulations demonstrating the influence of beam loading in a laser-plasma accelerator, for a laser with $a_0 = 3$. Each wakefield is loaded with a Gaussian bunch with different charge: 10 pC (left), 100 pC (middle), and 300 pC (right). The plots display the plasma density ($n_e/n_0 - 1$), with corresponding on-axis lineouts of the longitudinal electric field. As beam charge increases from left to right, the significant alteration of the wakefield structure becomes apparent.

wakefield. Generally, an acceleration length exceeding half the pump depletion length is challenging to achieve in practice.

1.2.5 Beam loading

Just as with the achievable energy, it would be unphysical to assume there's no limit to the amount of charge that can be accelerated. To obey the law of energy conservation, the electron bunch needs to absorb energy from the wakefield. The effect through which this happens is called *beam loading* [89, 90].

The concept of beam loading is hinged on the interplay between the electron beam and the plasma wave generated by the laser pulse. The space charge field of the accelerated electron beam can "load" the wakefield, thereby altering its structure. Particularly, when the beam's charge is sufficiently high, it can level out the wakefield region it inhabits.

The simulations presented in fig. 1.5 show this phenomenon. We can observe that for a low beam charge of 10 pC, the wakefield is barely affected, preserving its initial structure. As the beam charge is increased to 100 pC, the influence of beam loading becomes evident as the accelerating field is significantly flattened across the region occupied by the beam. This results in a less variable accelerating gradient across the beam, which can be beneficial for reducing energy spread. However, when the charge reaches 300 pC, the field reverses, such that the tail end of the bunch experiences a lower field than the head, indicating an overloading of the wake.

In an ideal beam loading situation, the wakefield can be adjusted to sustain a constant accelerating field across the electron bunch's length. This scenario leads to a low energy spread especially when accelerated to high energies. However, the realization of *optimal beam loading* is a complex task. It necessitates precise control over variables such as the electron beam's position and current profile.

The ideal beam loading scenario in the blowout regime would involve a triangular electron beam current profile [91]. In practice, the laser-plasma accelerator may operate in other regimes that aren't strictly the blowout. These can introduce variations in the wakefield structure and thus influence the optimal current profile for beam loading.

Additionally, over the course of propagation, the driver pulse itself undergoes evolution, which in turn leads to changes in the wakefield structure. This dynamic nature of both the driver and the wakefield needs to be factored into any optimization process for beam loading. Consequently, achieving optimal beam loading is a complex task that requires a detailed understanding of the interplay between the plasma, the driver laser, and the electron bunch, likely aided by a combination of theoretical modelling, numerical simulations, and careful experimental design.

2 Bayesian Optimization

This chapter will give an introduction to the mathematical concepts behind Bayesian optimization. Most readers of this thesis may find themselves on friendlier terrain with laser or plasma physics than Bayesian statistics, and truth be told, I found myself in the same boat prior to this venture. With that in mind, the following sections will strive to unravel this complex topic.

Section 2.1 serves as an entryway to the concepts of Bayesian statistics and the mathematical jargon used throughout this chapter, followed by an exploration of Gaussian process regression, the modeling foundation of Bayesian optimization, in section 2.2. The details of Bayesian optimization itself will be unfolded in section 2.3.

For those readers who have a date with destiny—or simply a tight schedule—or for those less intrigued by the intricacies of the Bayesian approach, a digestible summary in section 2.4, complemented by a quick reference glossary in table 2.1, should provide ample knowledge to follow the discussions in this thesis.

2.1 Bayesian statistics

Before introducing the components of Bayesian optimization, this section will briefly discuss the fundamental concepts of Bayesian statistics and is meant to mostly introduce the mathematical language that is used throughout this chapter. For a more comprehensive introductions into the topic, see e.g., references [92] and [93].

At its core, the most distinguishing feature of Bayesian statistics lies in its interpretation of probabilities as a degree of uncertainty. This viewpoint contrasts with frequentist approaches, where probabilities represent the ratio of outcomes to a particular event, assuming the event could be repeated an infinite number of times.

To deal with uncertainty, Bayesian statistics provides an elegant formalism of how the combination of an initial believe with further information leads to an updated uncertainty about an event. This principle is expressed in Bayes' rule:

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)} \tag{2.1}$$

The equation describes the probability of an event A given an observation B, $p(A|B)$, or the *posterior* of A given B. This depends on the *prior* $p(A)$ which expresses the initial believe about the probability of A. The prior is weighted with the *likelihood* $p(B|A)$ which measures the probability of the observation B in the case of event A. The prior and likelihood are normalized

with the *evidence* or *marginal likelihood* $p(B)$ which is the probability of observing B for any event. These rather fuzzy definitions shall be further motivated in the example below.

A common application of Bayes' principle is the interpretation of medical tests given a finite accuracy of the test result. An example for this, that has gained relevance during the writing of this thesis, are rapid tests that were used to fight the global COVID-19 pandemic. Confronted with a positive result a reasonable reaction is to question the validity of the test. Using Bayes' theorem we can calculate the posterior probability for this case, i.e., the probability of being infected given a positive test result $p(\text{infected}|+)$, from prior information that is available. For this example eq. (2.1) reads as

$$p(\text{infected}|+) = \frac{p(+|\text{infected}) \cdot p(\text{infected})}{p(+)}. \tag{2.2}$$

A reasonable prior for the probability of being infected, $p(\text{infected})$, is the local weekly infection rate, e.g., 0.1% of the population. The likelihood, $p(+|\text{infected})$, i.e., the probability that a test correctly shows a positive result, is given by the test's sensitivity of ∼99%. Finally, we need to find the evidence, $p(+)$, that a test shows a positive result. This is the combination of all results leading to a positive test result, so in addition to the sensitivity we need to know the specificity of the test. This is the probability of an accurate negative test, which is also ∼99%. More precisely, we need the opposite of this, i.e., the probability of a test showing a positive result for a healthy person. With this the evidence is

$$\begin{aligned} p(+) &= p(+|\text{infected}) \cdot p(\text{infected}) + p(+|\text{healthy}) \cdot p(\text{healthy}) \\ &= 0.99 \cdot 0.001 + 0.01 \cdot 0.999 \approx 0.011. \end{aligned} \tag{2.3}$$

So in combination, the probability that the positive test result actually indicates an infection is

$$p(\text{infected}|+) = \frac{0.99 \cdot 0.001}{0.011} \approx 9\% \tag{2.4}$$

which seems shockingly low, and in fact is a consequence of the low probability to be infected in the first place. However, the beauty of Bayesian statistics is that it provides a framework to incorporate additional information by updating our prior. So we can reduce the uncertainty of the result with a second independent test. As a prior we use the posterior probability after the first test. The probability of being infected given the second positive tests then increases to

$$\begin{aligned} p(\text{infected}|+_2) &= \frac{p(+|\text{infected}) \cdot p(\text{infected}|+)}{p(+_2)} \\ &= \frac{0.99 \cdot 0.09}{0.098} \approx 91\% \end{aligned} \tag{2.5}$$

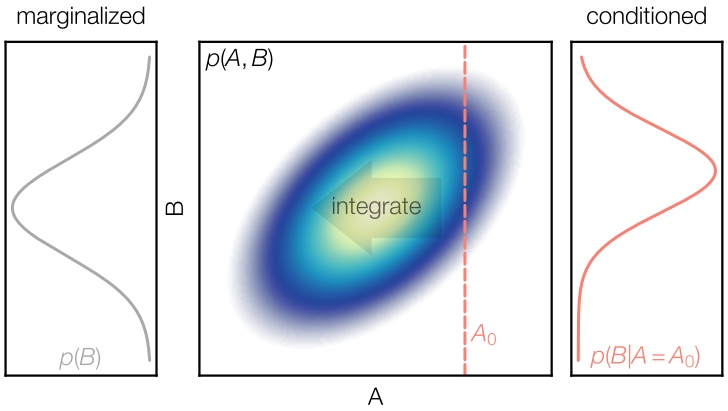

Figure 2.1 – Marginalization and conditioning on a joint probability density distribution. Integration of the joint distribution $p(A, B)$ over A gives the marginal distribution $p(B)$ (left). A slice through the joint distribution at A_0 represents the conditioned probability distribution $p(B|A = A_0)$.

2.1.1 Conditional probabilities

The example above implicitly used some concepts of probability that require a more detailed definition at this point. These are the *marginalization* $p(B)$ and *conditioning* $p(B|A)$ of a joint probability distribution $p(A, B)$. The marginalized distribution expresses the probability of a random variable independent of another. In the example above this is the chance that a test shows a positive result independent whether the patient has the disease or not. It is the sum of all properties leading to a given result

$$p(B) = \sum_A p(A, B). \tag{2.6}$$

For continuous random variables, which are more relevant in the following, the marginalization of $p(A, B)$ over A is

$$p(B) = \int_A p(A, B) dA. \tag{2.7}$$

The left-hand side of fig. 2.1 shows the marginal probability density function of a bivariate normal distribution that is generated by integrating along one of the variables.

The conditioning of a probability distribution can be understood as restricting the space of considered possibilities. In the example this is the probability of being infected given the positive test result. The conditional probability is defined as

$$p(B|A) = \frac{p(A, B)}{p(A)}. \tag{2.8}$$

The right-hand side of fig. 2.1 also shows a visual interpretation of the conditioned probability of B given that A has a certain value A_0. As shown $p(B|A = A_0)$ can be seen as a slice through the joint probability of A and B where the condition $A = A_0$ is fulfilled.

2.1.2 Regression

The context in which Bayesian statistics is relevant in this thesis is to make predictions about unobserved events based on previously measured data. The following gives an example, adapted from Ref. [94], where Bayesian reasoning is applied to a linear regression problem.

The goal of linear regression is to find the weights $\boldsymbol{w}$ in a linear model that describe a set of observations or in Bayesian terms, the probability distribution of the weights given the observed data. This requires the definition of a linear model that describes the relationship of the input $\boldsymbol{x}$ and $\boldsymbol{w}$ with the observation y:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{w}, \qquad y = f(\boldsymbol{x}) + \varepsilon, \tag{2.9}$$

where f is the function value that combined with some noise ε results in the observation[1]. Here the observation noise is assumed to follow a normal distribution with variance σ_ε^2 centered around zero

$$\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2). \tag{2.10}$$

In the following it is assumed that σ_ε is known, methods to infer this information from the data will be discussed later.

Applying Bayes' principle, the aim now is to find the posterior probability distribution of $\boldsymbol{w}$ given a set of n observations $p(\boldsymbol{w}|\boldsymbol{y}, X)$, where $\boldsymbol{y} = (y_i \mid i = 1, \dots, n)$ and $X = (\boldsymbol{x}_i \mid i = 1, \dots, n)$ are the single measurements as a stacked vector and matrix, respectively. According to eq. (2.1) the posterior over $\boldsymbol{w}$ is given by

$$p(\boldsymbol{w}|\boldsymbol{y}, X) = \frac{p(\boldsymbol{y}|X, \boldsymbol{w}) \cdot p(\boldsymbol{w})}{p(\boldsymbol{y}|X)}. \tag{2.11}$$

Recall that the choice of the prior $p(\boldsymbol{w})$ expresses the initial believe of probable values of $\boldsymbol{w}$ and is rather problem dependent. For mathematical simplicity, here it is chosen to also be a (multivariate) normal distribution centered around zero with covariance matrix Σ_p

$$\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \Sigma_p) := \frac{1}{\sqrt{(2\pi)^k \det \Sigma_p}} \exp\left(-\frac{1}{2}\boldsymbol{w}^\top \Sigma_p^{-1} \boldsymbol{w}\right), \tag{2.12}$$

where k is the dimensionality of $\boldsymbol{w}$. This prior expresses the believe that weights near zero are more likely.

The next term to determine is the likelihood $p(\boldsymbol{y}|X, \boldsymbol{w})$, which is the probability of the observation given the configuration of the model, or simply speaking how likely is it to observe

[1] A constant offset in $f(x)$ can be achieved by adding a dimension to $\boldsymbol{x}$ that is constant.

the given data if the model has certain weights. According to the model in eq. (2.9), y_i is the sum of the function value f and the Gaussian noise ε so the distribution of y_i is also a normal distribution that is centered around $f(\boldsymbol{x}_i)$. Considering the product of all n observations, this is

$$\begin{aligned} p(\boldsymbol{y}|X,\boldsymbol{w}) &= \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i,\boldsymbol{w}) \\ &= \frac{1}{(2\pi\sigma_\varepsilon^2)^{n/2}} \exp\left(-\frac{(\boldsymbol{y}-X^\top\boldsymbol{w})^2}{2\sigma_\varepsilon^2}\right) = \mathcal{N}(X^\top\boldsymbol{w},\sigma_\varepsilon^2 I_n), \end{aligned} \tag{2.13}$$

where I_n is the n-dimensional identity matrix. The traditional least squares method of linear regression is equivalent to maximizing this likelihood term.

The marginal likelihood $p(\boldsymbol{y}|X)$ is independent of $\boldsymbol{w}$ and can therefore be ignored at this point.

So finally, combining all $\boldsymbol{w}$ dependent terms in the prior and the likelihood leads to the posterior distribution (a detailed derivation can be found in Ref. [94])

$$p(\boldsymbol{w}|\boldsymbol{y},X) \sim \mathcal{N}(\sigma_\varepsilon^{-2}A^{-1}X\boldsymbol{y},A^{-1}), \tag{2.14}$$

with $A = \sigma_\varepsilon^{-2}XX^\top + \Sigma_p^{-1}$.

The top panels of fig. 2.2 show the prior and posterior distributions of $\boldsymbol{w}$ for the case of a one dimensional problem. The data samples are drawn according to eq. (2.9) with $\boldsymbol{w} = (w_0,w_1)^\top = (-1,0.5)^\top$ with $\sigma_\varepsilon = 0.5$. By incorporating the available data, the fairly unspecific prior narrows in on the more informed posterior distribution of $\boldsymbol{w}$. With this it is now possible to make a probabilistic statement about the underlying model that describes the data, e.g., "How likely is it that the data is independent of x?". According to the posterior distribution the probability for that, i.e., $p(w_1 = 0|\boldsymbol{y},X)$, is very close to zero.

Often when modeling data the aim is to make predictions about points that have not yet been observed. With most non-Bayesian methods the approach for this would be to select a $\boldsymbol{w}$ by some criterion, e.g., maximum likelihood (least squares regression) or maximum posterior (ridge regression), and fix the model with this value. However, these approaches lose the information about the uncertainty of the model. So the Bayesian approach is to yet again treat the prediction as a random variable that follows a probability density distribution, the *posterior predictive distribution*.

Therefore, the predictions about $f(\boldsymbol{x}_*)$ at the new positions x_* are expressed by marginalizing the distribution of possible outcomes at x_* over $\boldsymbol{w}$, i.e., averaging over all possible model configurations weighted by their posterior likelihood

$$\begin{aligned} p(f(\boldsymbol{x}_*)|\boldsymbol{x}_*,X,\boldsymbol{y}) &= \int p(f(\boldsymbol{x}_*)|\boldsymbol{x}_*,\boldsymbol{w})p(\boldsymbol{w}|\boldsymbol{y},X)d\boldsymbol{w} \\ &= \mathcal{N}(\sigma_\varepsilon^{-2}\boldsymbol{x}_*^\top A^{-1}X\boldsymbol{y},\boldsymbol{x}_*^\top A^{-1}\boldsymbol{x}_*). \end{aligned} \tag{2.15}$$

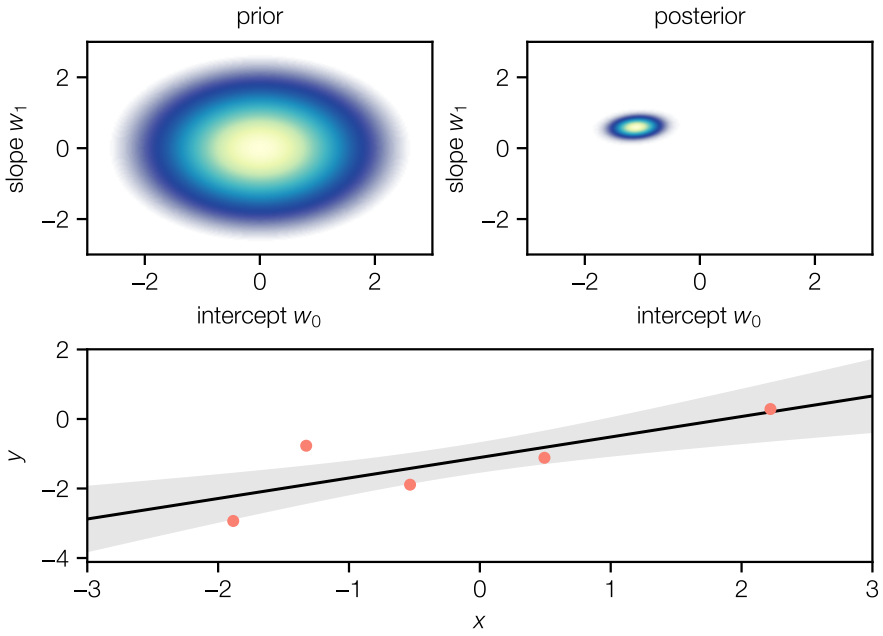

Figure 2.2 – Bayesian linear regression with noisy data samples (red dots). The prior distribution (upper left) expresses the believe that the model parameters w_0 and w_1 are close to zero. Conditioning the model on the data yields the posterior distribution (upper right) according to which the data most likely follows a model with finite slope and intercept. Bottom: Mean and confidence ($\pm$ two standard deviations) of the posterior predictive distribution (eq. (2.15)).

The lower panel in fig. 2.2 shows the mean and $\pm$ two standard deviations of the posterior predictive distribution. Instead of a single prediction that is most certain by a given criterion, linear regression in the Bayesian sense allows to make a statement about the realm of possible outcomes given the fitted data.

2.2 Gaussian process regression

Although linear models like the one described above are important tools to describe many aspects of physics including laser-plasma accelerators, they are often not expressive enough for the non-linear dynamics that dominate the laser-plasma interaction. One approach to tackle this is to use higher order parametric models. In this context, *parametric* means that the model is fully described by its parameters like the weights in the case above. Parametric models that capture more complex behavior are, for example, higher order polynomials or, on the more complex end of the spectrum, neural networks.

A different approach are *non-parametric* models that are defined by the data itself. One prominent example for this kind of model is a simple spline fit for which there is no global model of parameters but just local interpolations between data points. Non strictly speaking, this increases the value of each data point available to the model and consequently non-parametric models tend to produce good results also with a sparse amount of information.

The class of non-parametric models that is commonly used for Bayesian optimization are so-called Gaussian processes which will be discussed in this section. The discussion here closely follows the excellent introduction to Gaussian process regression by Rasmussen and Williams [94] and will cover the relevant aspects of Gaussian processes in the context of Bayesian optimization.

2.2.1 Definition

A Gaussian process is the generalization of a multivariate normal distribution for the case of infinitely many random variables. Additionally, per definition each finite subset of these infinitely many random variable again follows a multivariate normal distribution.

To understand the implications of this rather abstract concept, consider a simple normal distribution in one dimension. The distribution describes the realm of possible outcomes for a given event, e.g., a measurement of some quantity at the input value x_0. Consequently, each measurement is a sample of this distribution and returns a single value y_0. Next, take a measurement of a set of inputs at (x_0, x_1), since the outcomes of these are potentially correlated they follow a joint normal probability distribution, which for each measurement gives a pair of samples (y_0, y_1). If now, instead of at discrete positions the measurement is performed at all possible locations along the entire domain x, the result of each measurement will be an infinite set of values $y(x)$. With each new measurement, the result will be a new $y(x)$ which then, in fact, is distributed according to a Gaussian process. Consequently, Gaussian processes describe the probability distribution over functions.

The fact, that each finite set of random variables follows a normal distribution is important here. When making statements about a Gaussian process, e.g., predictions, for a finite number of points it's a matter of dealing with normal distributions which as shown in the last section is often analytically tractable.

Although, a Gaussian process is admittedly not a concept that is particularly easy to grasp, it will prove to be very useful to build rich models around measured data and importantly to quantify uncertainty in these models.

Analogous to the normal distribution a Gaussian process is defined by a mean and a covariance. However, since the Gaussian process describes continuous variables both are functions as well. These relate to the collection of functions $f(\boldsymbol{x})$ according to

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})], \qquad \text{mean function,} \tag{2.16}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))], \qquad \text{covariance function,} \tag{2.17}$$

where $\mathbb{E}$ denotes the expected value. With this, a common notation to say that a function $f(\boldsymbol{x})$ follows a certain Gaussian process is,

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')). \tag{2.18}$$

2.2.2 Kernel functions

Although Gaussian processes define a probability distribution over functions, they don't consist of all possible functions but are confined to functions of a specific characteristic. These are fully defined by the covariance function of the Gaussian process, here also referred to as kernel functions. The covariance is a measure for the similarity between the outcomes $f(\boldsymbol{x})$ and $f(\boldsymbol{x}')$ for the inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$. Simply speaking, if the covariance for two inputs is large they tend to produce similar outputs. The choice of the covariance function in a Gaussian process is arbitrary and problem dependent. Some of the most commonly used kernels and their characteristics will be discussed in the following.

One of the simplest and widely found kernel functions is the *squared exponential* covariance function, more commonly known as *radial basis function (RBF)* kernel

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right). \tag{2.19}$$

With this kernel the similarity between two points falls off as a Gaussian function of their distance with the characteristic lengthscale l.

To get an intuition what this means for the resulting class of functions that the Gaussian process describes, it makes sense to draw some samples f_* from it at some test positions X_* and plot them. Since we are looking now only at a finite number of points, the Gaussian process reduces to a simple normal distribution,

$$f_* \sim \mathcal{N}(\boldsymbol{m}(X_*), K(X_*, X_*)), \tag{2.20}$$

where $\boldsymbol{m}$ and K are the vector and matrix that result when evaluating the mean and covariance functions over all possible combinations of the test points. Figure 2.3 shows some samples

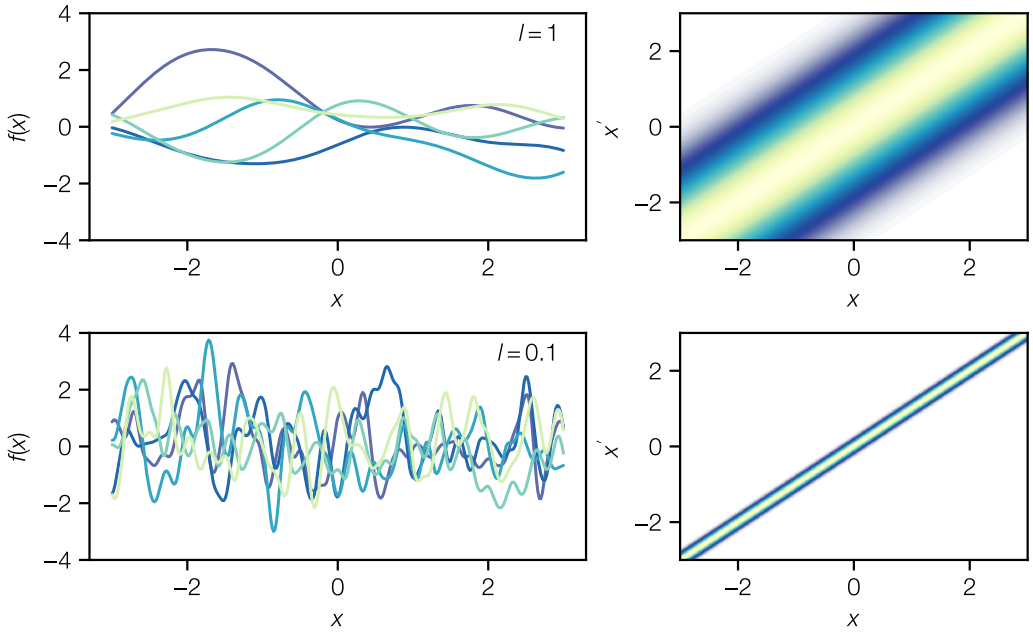

Figure 2.3 – Gaussian process with RBF covariance function. Left column: Samples drawn from Gaussian processes with large (top, $l = 1$) and small (bottom, $l = 0.1$) lengthscale parameters. Right column: Corresponding covariance matrices for the plotted x values in the left column (blue to yellow, white indicates zero covariance).

from a Gaussian process with an RBF kernel and the corresponding covariance matrix (the mean function is set to zero in this example).

As shown, the lengthscale of the kernel determines the rate at which $f(x)$ changes as a function of x. This can be understood when comparing the covariance matrices of the two cases. With the long lengthscale there is a finite covariance over the entire range which means that points far from each other still are correlated and therefore tend to have similar values. In the case of a small lengthscale, the covariance is much more confined around the diagonal which means that the range of influence of one point is confined to its close proximity. Points further away are fully uncorrelated and therefore the functions can vary on a shorter lengthscale. Nevertheless, both cases lead to smooth functions and in fact the RBF kernel produces only functions that are infinitely differentiable.

For some use cases the smoothness of the RBF kernel can be problematic. The *Matérn* kernel [95] is a generalization of the RBF kernel that allows to control the smoothness, i.e., the differentiability, of the contained functions. It is defined as

$$k(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{\Gamma(\nu)2^{\nu-1}}\left(\frac{\sqrt{2\nu}}{l}|\boldsymbol{x}-\boldsymbol{x}'|\right)^{\nu} K_\nu\left(\frac{\sqrt{2\nu}}{l}|\boldsymbol{x}-\boldsymbol{x}'|\right), \tag{2.21}$$

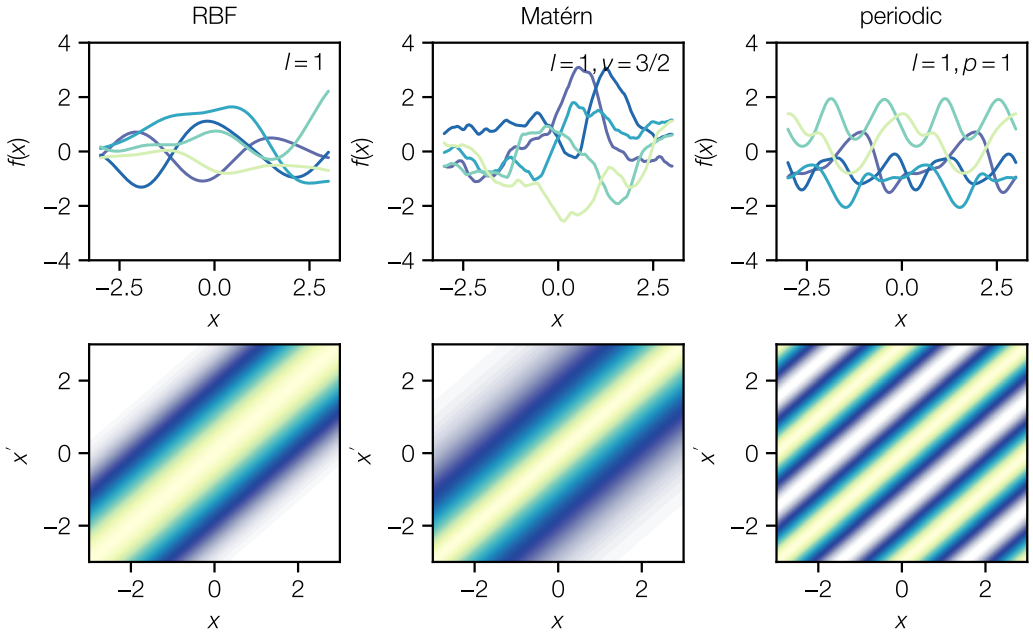

Figure 2.4 – Characteristics of different kernel functions. Samples from Gaussian processes with given kernels (top) and their corresponding covariance matrices (bottom).

where K_ν is a modified Bessel function and Γ it the gamma function. The Matérn kernel has two parameters: again the lengthscale parameter l, and the parameter ν that controls the smoothness of the considered functions. For $\nu \to \infty$ the Matérn kernel converges to the RBF kernel and for $\nu = 3/2$ and $\nu = 5/2$ it represents once and twice differentiable functions, respectively. Figure 2.4 shows a comparison between samples from Gaussian processes using an RBF and a Matérn kernel. Even though, the covariance matrices for the two cases look very similar, the sampled functions look distinctly different. By not enforcing infinite differentiability, the Matérn kernel produces less smooth functions.

The right panel in fig. 2.4 also shows how kernels can also enforce very distinct characteristics on Gaussian processes. The kernel that is used here is periodic and consequently creates periodic patterns in the sampled functions. The periodic kernel is given by:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{2\sin^2(\pi|\boldsymbol{x} - \boldsymbol{x}'|/p)}{l^2}\right), \tag{2.22}$$

with the parameters l and p that define the lengthscale and periodicity, respectively. Note how the samples are not just simple sinusoids so they can express more complex periodic behaviors. This kernel for instance can be used to model recurrent features in a time series, like oscillations of a system or seasonal effects.

These three kernels are just some frequently encountered examples and in fact there is a vast amount of different covariance functions to describe all types of functional behavior (see

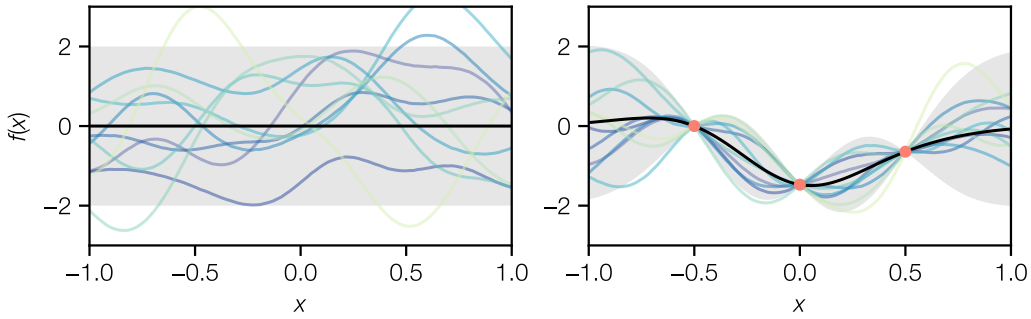

Figure 2.5 – Samples (colored lines) from a Gaussian process prior (left panel) and posterior (right panel) after conditioning on data (red dots). The black line indicates the mean of the Gaussian process and the gray area its 2σ confidence region. By conditioning the process its mean and confidence collapse so that the samples are forced through the data points which in this example are assumed to be free of noise.

chapter 5 of [94] for a detailed breakdown). This highlights the potential expressiveness that Gaussian processes can have compared to simple linear models.

2.2.3 Conditioning

So far these Gaussian processes are merely a mathematical construct to describe distributions over functions and are not much use to model real world data. But again recall that evaluated on a finite number of points, and this is typically what is of interest when building models, a Gaussian process reduces to a multivariate normal distribution. As shown in the previous section, there is a simple way to incorporate knowledge in such a distribution through conditioning and marginalization with Bayes' theorem.

To build a model using a Gaussian process prior and some training data $X, \boldsymbol{f} = \{\boldsymbol{x}_i, f_i | i = 1 \dots n\}$ and to make predictions $\boldsymbol{f}_*$ at some test inputs X_* the joint distribution of $\boldsymbol{f}$ and $\boldsymbol{f}_*$ needs to be considered. Similar to eq. (2.20) this is given by

$$\begin{bmatrix} \boldsymbol{f} \\ \boldsymbol{f}_* \end{bmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(X,X) & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right). \tag{2.23}$$

Here, again $K(X,X)$ refers to the matrix that results from evaluating the covariance function over all combinations of inputs in X.

The posterior distribution, i.e., the distribution over f_* that are compatible with the training data, is now obtained by conditioning the joint distribution eq. (2.23) on the training data. Making use of the properties of a normal distribution this can be done analytically [94] and the posterior distribution then is

$$\begin{aligned} \boldsymbol{f}_* | X_*, X, \boldsymbol{f} \sim \mathcal{N}(&K(X_*,X)K(X,X)^{-1}\boldsymbol{f}, \\ &K(X_*,X_*) - K(X_*,X)K(X,X)^{-1}K(X,X_*)). \end{aligned} \tag{2.24}$$

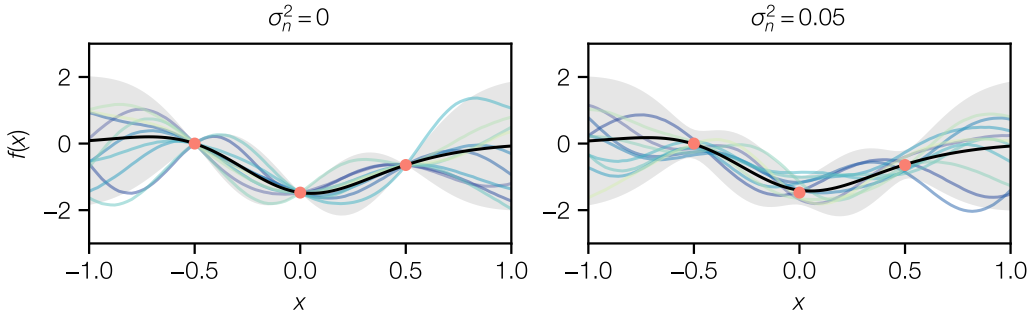

Figure 2.6 – Predictions of a Gaussian process assuming noiseless (left) and noisy (right) data. Both cases use the same data points for regression but in the right case they assumed to be noisy observations of the underlying process with variance 0.05. Consequently, the mean of the predictive distribution does not cross the data exactly and there remains a finite uncertainty around the observations.

Figure 2.5 shows an example of conditioning a Gaussian process on data points. The prior, i.e., the process before conditioning, has a constant mean and variance. After conditioning, the samples of the distribution are forced through the data and consequently so is the mean. Since the samples no longer vary around the data points the variance collapses at these positions.

In terms of a more practical interpretation, close to the data points the confidence in the model prediction, i.e., the mean of the Gaussian process, is high and the further a test point gets from a known training point the more uncertain the prediction of the model gets.

Obviously, for modeling real world experimental data, the model in fig. 2.5 is too confident in the data points. Since there is observation noise on the data the mean of the underlying process does not necessarily coincide with all data points.

The presence of observation noise can easily be expressed in the covariance function of the Gaussian process. As in the example of linear regression in the last section, the training data y is not sampled from $f(\boldsymbol{x})$ but contains additional noise ε that is assumed to be normally distributed with variance σ_n^2, i.e., $y = f(\boldsymbol{x}) + \varepsilon$. In this case the covariance function over the training data is just

$$cov(y_i, y_j) = k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sigma_n^2 \delta_{ij} \tag{2.25}$$

where δ_{ij} denotes the Kronecker delta which is one if $i = j$ and zero otherwise. In analogy to eq. (2.24) the posterior distribution for $\boldsymbol{f}_*$ given noisy observations $\boldsymbol{y}$ is

$$\begin{aligned} \boldsymbol{f}_* | X_*, X, \boldsymbol{y} \sim \mathcal{N}(&K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} \boldsymbol{y}, \\ &K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*)), \end{aligned} \tag{2.26}$$

where I is the identity matrix.

Figure 2.6 shows the mean and standard deviations of two Gaussian processes conditioned on the same data, but in one case assuming it to be the underlying process directly and in the other to be noisy samples of it. In the noisy case the sampled functions no longer need to pass

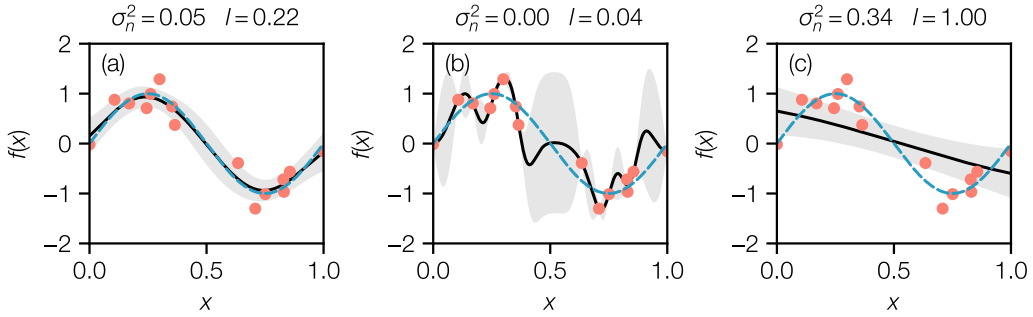

Figure 2.7 – Gaussian process models (RBF kernel) with different hyperparameters conditioned on the same data (red dots). Depending on the choice of lengthscale l and noise σ_n^2 of the covariance function the model is able to generalize the underlying sine function (blue) better or worse. Without noise (middle) the model requires a small lengthscale to account for all data points. This leads to erratic behavior of the mean when interpolating between the measured data. If the lengthscale is too large (right), the information in the data is interpreted as noise leading to a too simple model.

through the observations directly. Consequently, also the mean of the distribution does no longer exactly cross all data points and there remains a finite uncertainty around them.

2.2.4 Hyperparameters

At the beginning of this section, it was mentioned that Gaussian processes are non-parametric models. A given model is fully described by its mean and covariance function, which are determined by conditioning the Gaussian process prior on the data. However, the choice of covariance function and its parameters, such as the lengthscale or noise level, can lead to different expressions of the Gaussian process. The selection of these parameters determines how well a Gaussian process prior can adapt to data. As the parameters of the covariance function are not parameters in the same sense of weights in a linear model, they are commonly referred to as *hyperparameters*.

Figure 2.7 shows Gaussian process models using RBF kernels with different noise and lengthscales that are conditioned on data that was sampled from a sine function with added noise. In the right panel of the figure, the process with the correct noise and lengthscale generalizes the underlying function very well and is accurate in regions without observations. When assuming no noise on the data, as depicted in the center panel, the mean is forced through all data points and consequently the process requires a short lengthscale. This has the effect that the model varies rapidly away from the observations, leading to wrong predictions. This effect is often referred to as overfitting, i.e., specializing on the training data and thereby losing the ability to generalize. In the other extreme when assuming a very long lengthscale, all variations of the data need to be described by the noise of the model. As depicted in the right panel of fig. 2.7, this leads to a too simple model that does not accurately describe the underlying behavior of the data.

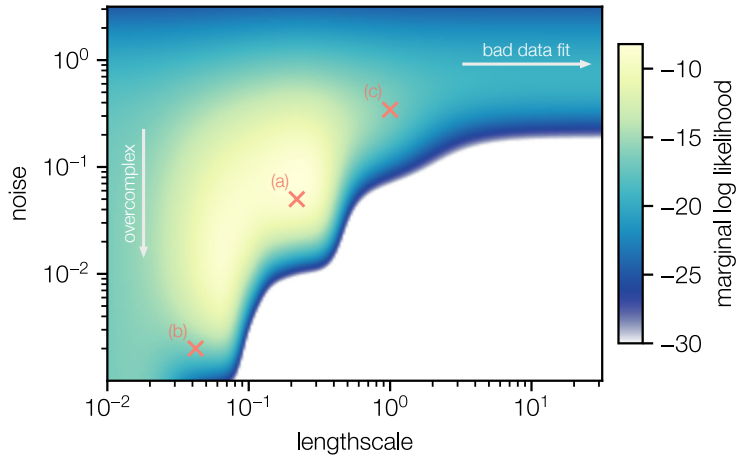

Figure 2.8 – Log marginal likelihood for data and example models from fig. 2.7 (crosses). Complex models with small lengthscales tend to overfit on the data. On the other end with too simple models with large lengthscales much information in the data is interpreted as noise. The marginal likelihood balances the data fit and model complexity.

In many machine learning approaches, a common method to find proper model parameters is to define a loss function that quantifies the prediction error of the model with regard to some test data. This loss function could for example be the mean absolute error, i.e., the mean difference between model prediction and the test data. The parameters of the model can then be optimized for minimal loss.

With probabilistic models like Gaussian processes there is a natural quantity that describes the probability of measuring some data given a certain underlying model (or model hyperparameters), the *marginal likelihood*.

Since the marginal likelihood can vary drastically in amplitude depending on the model, typically its logarithm is considered for numerical stability. For a Gaussian process with hyperparameters θ the log marginal likelihood for a set of n noisy observations $(X, \boldsymbol{y})$ is given by [94]

$$\log p(\boldsymbol{y}|X, \theta) = -\frac{1}{2}\boldsymbol{y}^\top [K + \sigma_n^2 I]^{-1}\boldsymbol{y} - \frac{1}{2}\log|K + \sigma_n^2 I| - \frac{n}{2}\log 2\pi, \tag{2.27}$$

where K is the covariance matrix that results from evaluating the covariance function for all inputs. The first two terms of this equation have distinct meanings in describing the quality of the model. The first term describes *how far off* the data is from the distribution described by the Gaussian process. If $\boldsymbol{y}$ is equal to the mean of the Gaussian process at the respective positions this term becomes zero. The second term describes the complexity of the model. The determinant of the covariance matrix can be understood as a measure for the dependency between the different data points. If it is zero, all y are perfectly correlated and therefore the resulting functions need to be constant. On the contrary if it is one, all points are unrelated and the function are very complex and can adapt to arbitrary data. Note that this term is subtracted which means that simple models, e.g., slowly varying ones are favoured. This is

resembling the principle of Occam's razor that simpler explanations for an observation are more likely.

In combination, the log marginal likelihood favours models that fit the data well but penalizes them when they are overly complex. Figure 2.8 shows the marginal log likelihood for the example data from fig. 2.7 as a function of the model's lengthscale and noise level. The marginal likelihood has a maximum for at the parameters that reproduce the underlying sine function as shown in fig. 2.7. In fact the parameters of fig. 2.7 were found by maximizing the marginal log likelihood. When increasing the lengthscale, which reduces the complexity, the data fits the model worse and needs to be described by higher noise. On the other end, when reducing the noise the data fit can be perfect at the cost of high model complexity through the short lengthscale. When increasing the lengthscale while decreasing the noise the model does not describe the data and quickly falls off.

Typically, the more data is available the more pronounced the optimum of the marginal likelihood gets.

Also the derivatives of the log marginal likelihood can be computed without significantly increasing the computational cost [94]. With them, conventional gradient based optimizers can be used to find the hyperparameters $\hat{\theta}$ that maximize the log marginal likelihood and therefore have the highest probability of describing the data. It is also possible to consider a prior on the hyperparameters θ for example if reasonable knowledge about the lengthscale of the problem is available a priori. In that case the product of the marginal likelihood and the prior of the hyperparameters $p(\theta)$ is maximized, or equivalently the sum of their logarithms [92]

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}(\log p(\boldsymbol{y}|X, \theta) + \log p(\theta)) \tag{2.28}$$

2.3 Bayesian optimization

Bayesian optimization [96] is a strategy to find global extrema in noisy black-box functions that are costly to evaluate. In this context the term black-box describes the circumstance that a priori there is no knowledge about the topology of the problem and its derivatives. This includes, for example, non-convex problems that can have many local optima. Therefore, information about the problem can only be gained by costly evaluations. The cost hereby typically refers to the time to get a solution but can also refer to monetary or computational cost for example when running simulations. Additionally, in experimental applications the underlying black-box function cannot be accessed directly but only via noisy samples.

The combination of substantial cost evaluating the function, that is being optimized, and noise on the observations limits the applicability of conventional optimization algorithms like Newton methods. The approach of Bayesian optimization to overcome these issues is to employ a *surrogate model* which approximates the actual optimization problem and is cheap to evaluate and free of noise. Consequently, this surrogate model then again allows to use conventional optimization methods. Due to their ability to cope with noise and little available data, Gaussian process models are typically used for this task.

However, a model that is based on only a few measurements is bound to have inaccuracies and therefore does not reveal the true optimum right away. Therefore, Bayesian optimization employs a strategy to choose the parameters of additional measurements of the black-box function that are then used to improve the model. This search heuristic is the so-called *acquisition function*.

2.3.1 Acquisition function

The acquisition function quantifies the utility of a possible next measurement in finding the global optimum. There are various different acquisition functions but most of them share the same approach of what is often called balancing exploitation and exploration of the parameter space. This means that the acquisition function identifies either regions of the parameter space where evaluations with good results have already been made to improve upon these (exploitation), or regions where data is still sparse that could therefore contain a yet unknown optimum (exploration). Here, the concept of Gaussian processes proves useful since it directly provides a measure for (un)certainty with the variance of the posterior predictive distribution given by eq. (2.26).

A common and simple acquisition function is *expected improvement* (EI) [96]. It bases its definition on the best observation made so far, denoted as f^*. The function expresses the anticipated improvement of the optimized function over f^* for all possible input values $\boldsymbol{x}$, based on the surrogate model that is conditioned on existing measurements. In simpler terms, the EI function forecasts how much better we can do over our current best value for each input $\boldsymbol{x}$.

This is mathematically represented as

$$\mathrm{EI}(\boldsymbol{x}) := \mathbb{E}[\max(f(\boldsymbol{x}) - f^*), 0)], \tag{2.29}$$

where $\mathbb{E}$ symbolizes the expected value. The term $\max(f(\boldsymbol{x}) - f^*), 0)$ essentially indicates that there is no such thing as negative improvement. Instead, when $f(\boldsymbol{x}) < f^*$', it results in zero improvement, i.e., no betterment is gained compared to our current best f^*.

With the Gaussian process model $f(\boldsymbol{x})$, eq. (2.29) can be solved analytically [96, 97] to give the expression

$$\mathrm{EI}(\boldsymbol{x}) = \sigma(\boldsymbol{x}) \left[\phi(u(\boldsymbol{x})) + u(\boldsymbol{x})\Phi(u(\boldsymbol{x}))\right], \tag{2.30}$$

where $u(\boldsymbol{x}) = (m(\boldsymbol{x}) - f^*)/\sigma(\boldsymbol{x})$ with the mean $m(\boldsymbol{x})$ and variance $\sigma(\boldsymbol{x})^2 = K(\boldsymbol{x}, \boldsymbol{x})$ of the surrogate model and ϕ and Φ denote the probability density function and the cumulative distribution function of the normal distribution, respectively.

The point that gives the largest expected improvement is used as the input for the next measurement. To find it, there is a similar expression for the derivative of the expected improvement which can be used to efficiently find the maximum of the acquisition function with gradient based (quasi-)Newton methods, e.g., the commonly used L-BGFS-B algorithm [98].

2.3.2 Optimization

Algorithm 1 Bayesian optimization of $y = f(\boldsymbol{x})$ over N iterations

```
Define Gaussian process surrogate model 𝓜
Define acquisition function α(x, 𝓜)
Make n_0 initial observations 𝒟 = (x_i, y_i | i = 0 ... n_0 )
n = n_0
while n < N do
    Update 𝓜 with 𝒟
    Maximize log marginal likelihood to find hyperparameters of 𝓜
    Find x_{n+1} = argmax α(x)
    Observe y_{n+1} = f(x_{n+1}), append dataset 𝒟 = (x_i, y_i | i = 0 ... n + 1 )
    n = n + 1
end while
Return best point
```

Now with a surrogate model and a search heuristic in place the optimization routine consists of iteratively maximizing the acquisition function to find the most promising input parameters, evaluating the black-box function at this position and then update the surrogate model with the result as described in algorithm 1. Given enough iterations the algorithm then efficiently explores the available parameter space to find the global optimum of the black-box function [99].

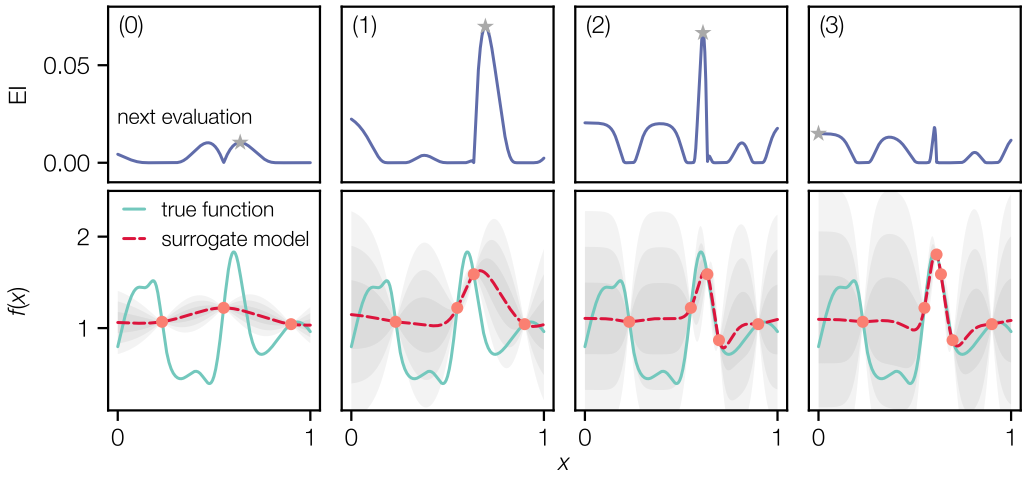

Figure 2.9 – One-dimensional Bayesian optimization with expected improvement as the acquisition function (upper row). The lower panels show the unknown ground truth (green) that is approximated with a Gaussian process model (red, 1,2,3-σ uncertainty in grey shaded areas) based on previous observations (red dots). In each iteration a measurement is made at the position where the acquisition function is largest (grey star).

Figure 2.9 depicts an example of using Bayesian optimization on a one-dimensional problem. After the first three initial observations (0) the surrogate model gives a very inaccurate representation of the real problem and in particular seems to assume a too long lengthscale to describe the true function. Still, taking into account the information that is present in the form of the model mean and that which is absent, i.e., the model variance, the acquisition function predicts improvements around the central data point. In the next step (1) the evaluation at this position in fact yields improvement. The acquisition function again shows the highest expected improvement close to the last evaluation. Although the following iteration (2) gives no improvement it brings important information about the lengthscale of the problem so that the surrogate now represents the true function much better in the relevant region. In the final step (3) the true maximum of the function is found.

2.3.3 Observation noise

In many real world optimization problems there is no direct access to the underlying black-box function, but evaluations are obscured by measurement noise. As discussed in section 2.2, Gaussian process models have a natural way of taking this observation noise into account. However, the simple definition of expected improvement in eq. (2.29) can be problematic in this case since the real noise-free value of the best observation so far is unknown. If, for instance, the best observation is a large outlier, this can falsely reduce the expected improvement relative to this outlier in the rest of the search domain. As shown in fig. 2.10 this can result in situations where the optimizer ignores the real global optimum which is lower than the outlier that was measured earlier.

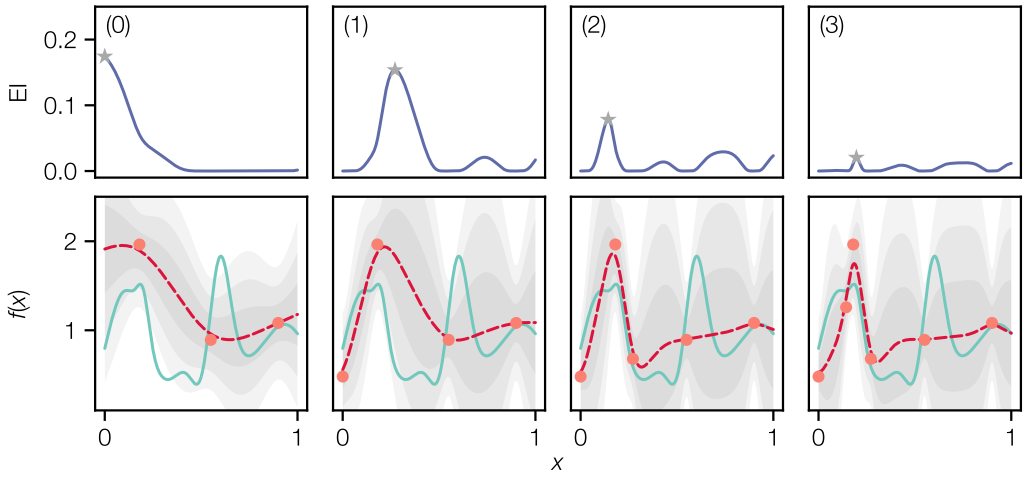

Figure 2.10 – Same optimization case as fig. 2.9 but with observation variance of 0.05. The outlier in the first step is interpreted as a very large best observation. This drastically reduces the expected improvement in the rest of the search domain and the optimizer ignores the true optimum.

Different approaches have been proposed to address this issue. A simple acquisition function that does not rely on any explicit previous observations but just on the surrogate model is *upper confidence bound* [100] which is simply the weighted sum of the model mean $m(\boldsymbol{x})$ and variance $\sigma(\boldsymbol{x})^2$

$$\mathrm{UCB}(\boldsymbol{x}) := f(\boldsymbol{x}) + \sqrt{\beta}\sigma(\boldsymbol{x}). \tag{2.31}$$

β is a parameter that balances the amount of exploration and exploitation, i.e., a large beta means that the acquisition function is large in regions with high uncertainty and consequently the optimizer is encouraged to explore the parameter space. Upper confidence bounds has been successfully used to optimize accelerators [101].

Another acquisition function that was shown to cope well with observation noise is *knowledge gradient* [102]. This method is based on the principle that the decision-maker is willing to accept a configuration that has not been evaluated previously and therefore trusts the surrogate model. The goal of knowledge gradient then is to find the point that if evaluated would give the largest expected improvement to the optimum of the surrogate model in the next step, i.e.,

$$\mathrm{KG}(\boldsymbol{x}) := \mathbb{E}[m^*_{n+1} - m^*_n | \boldsymbol{x}_{n+1} = n], \tag{2.32}$$

where m^*_n is the current maximum of the surrogate model and m^*_{n+1} that of the next iteration. Finding the maximum of the knowledge gradient involves solving a nested optimization problem and is therefore associated with high computational costs. Recent implementations that rely on Monte-Carlo sampling to approximate the integrals connected to the expected value in eq. (2.32) [103] reduce this costs, but solving knowledge gradient still remains significantly more costly than upper confident bound or expected improvement.

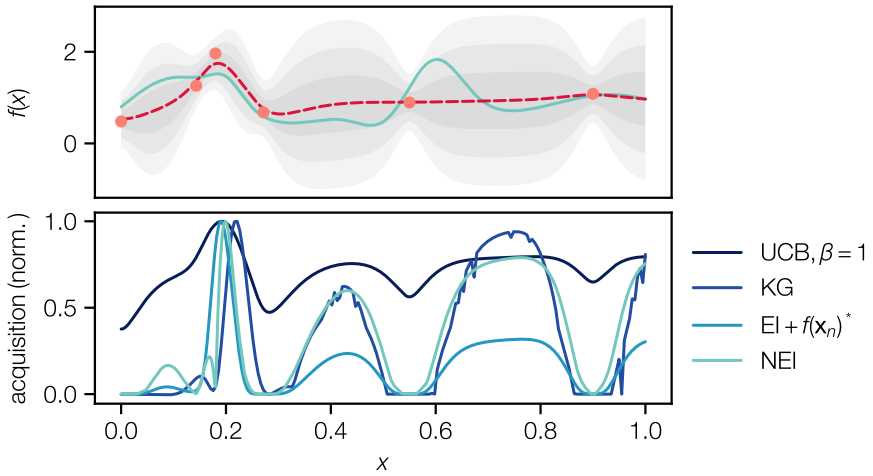

Figure 2.11 – Comparison of acquisition functions using the data from the last step in fig. 2.10. The rough features of knowledge gradient are caused by finite sampling when solving eq. (2.32). All acquisition functions are computed using the Bayesian optimization framework botorch [103].

Several adaptations to expected improvement have also been proposed, the simplest of which is to replace the best observation f^* in eq. (2.29) with the maximum of the values predicted by the surrogate model at the observed positions, i.e., $f(\boldsymbol{x}_n)^* = \max f(\boldsymbol{x}_n)$. With this a single observation can no longer compromise the whole optimization. Even though the model will likely overestimate the true function initially after measuring an outlier, it can be cured by additional observations. Still, if the observation noise is large this approach can lead to clustering of observations around single points and give non-ideal convergence rates [104].

A different solution is to treat the expected improvement acquisition in a more Bayesian way where the best observation so far is considered a random variable just like the optimization problem itself. Therefore, the best observation is defined by a probability distribution that is conditioned on the previous observations. Averaging of the original expected improvement function over previous observations then gives the *noisy expected improvement* acquisition function [105]

$$\mathrm{NEI}(\boldsymbol{x}) := \int_{\boldsymbol{f}^n} \mathrm{EI}(\boldsymbol{x}|\boldsymbol{f}^n) p(\boldsymbol{f}^n) d\boldsymbol{f}^n, \tag{2.33}$$

where $\boldsymbol{f}^n$ are the posterior distributions of the previous observations given their noisy observations, i.e., normal distributions according the Gaussian process model at the positions that have been observed. Again this acquisition function has no analytic expression, but can be solved efficiently and accurate with Monte-Carlo methods [105, 103].

Figure 2.11 shows a comparison of the described acquisition functions applied on the data from the last iteration of fig. 2.10. All methods would still recommend a next evaluation in the region where the outlier was initial measured, however knowledge gradient and noisy expected improvement have almost equally large values in other regions of the parameter space suggesting that exploration would still be valuable despite the large outlier. The modified

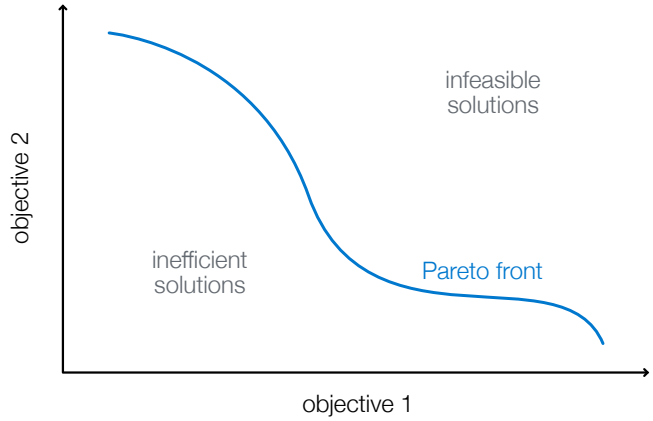

Figure 2.12 – Pareto front of two competing objectives assuming a maximization problem. The Pareto front separates the set of inefficient solutions that are suboptimal in both objectives from outcomes that are infeasible. The solutions on the Pareto front can only be improved in one objective by deteriorating the other.

expected improvement and upper confidence bound have far clearer preference for the region that has already been explored. This plot is just meant to give an example for the characteristics of the different methods and arguably doesn't help much in predicting the optimization behavior when using the different methods. Therefore, it is always helpful to benchmark different methods and see what works best for the specific problem at hand.

2.3.4 Multi-objective Bayesian optimization

Many optimization problems relevant in scientific and engineering applications involve multiple potentially competing objectives. In the context of particle accelerators, this could be for instance the maximization of the beam current while minimizing the energy bandwidth. Usually, such problems do not have a unique optimal solution but rather a range of desirable outcomes that involve various tradeoffs between the different objectives.

A common and simple approach in these scenarios is to scalarize the objectives by adding or multiplying them. The scalar objective can then be optimized as discussed above. However, the scalarization of the objectives has implications on the outcome of the optimization. By weighting the single objectives the scalarization inhibits a certain bias. Therefore, finding the scalarization that gives the desired results requires domain knowledge and an active decision on the desired result. For example, a decision-maker could decide that the energy spread of a beam is more important than its current and therefore give it a larger weight in the scalarization of the objectives. This specific domain knowledge might not always be present, when the interplay between the single objectives is complex or desired when there is no clear preference for a certain weighting.

An alternative approach to scalarizing the objectives is to search solutions at the so-called *Pareto front*. This refers to the hypersurface where bettering one objective inevitably results in the compromise of another, as illustrated in fig. 2.12. As per this definition, the Pareto front provides the ideal balance between the objectives. Thus, every point on this front symbolizes the optimum of a specific scalarization, showcasing a distinct equilibrium between the objectives.

Bayesian optimization can be applied to identify the best, or Pareto optimal, solutions in a multi-objective environment. This is achieved by using specialized acquisition functions and multi-output Gaussian processes. In the simplest scenario, these processes can be understood as independent models for each objective.

An intuitive extension of the expected improvement acquisition function is the *Expected Hypervolume Improvement* (EHVI) [106]. Within this framework, existing measurements are classified into *dominated* and *non-dominated* categories. A measurement is deemed dominated if another superior measurement exists across all objectives. Ideally, the non-dominated measurements would represent a subset of the Pareto front.

The hypervolume is defined as the volume encompassed by the non-dominated measurements and a reference point (see fig. 2.13). This reference point is usually the limit of acceptable outcomes. With this in mind, the hypervolume offers a measure of the quality of the set of non-dominated measurements. Larger hypervolumes indicate solutions closer to the true Pareto front.

The hypervolume is typically represented as a collection of n-dimensional rectangles, each spanned by a non-dominated point and the reference point. Consequently, the discovery of a new non-dominated point would lead to an improvement in the hypervolume, as depicted in fig. 2.13.

By employing the Gaussian process surrogate model it is possible to calculate the expected hypervolume improvement for all possible input parameters and therefore find the inputs that have the largest potential increase of the hypervolume

$$\mathrm{EHVI}(\boldsymbol{x}) := \mathbb{E}[\mathrm{HVI}(\boldsymbol{f}(\boldsymbol{x}))], \tag{2.34}$$

where HVI is the hypervolume improvement as depicted in fig. 2.13 and $\boldsymbol{f}(\boldsymbol{x})$ is the Gaussian process surrogate model. Exact solutions to eq. (2.34) exist [107–109] but scale exponentially with the number of objectives and can therefore be extremely costly to evaluate. Better performance can be achieved by approximating the expected value in eq. (2.34) with Monte-Carlo methods [110]. This further allows to define EHVI for multiple parallel evaluations and in the presence of observation noise [111].

A hand full of other methods for Bayesian optimization in multi-objective scenarios exist [96, 112, 113]. However, they either involve high computational complexity or achieve inferior performance compared to EHVI on synthetic problems. As a result, these alternative methods will not be extensively discussed or employed within the scope of this thesis.

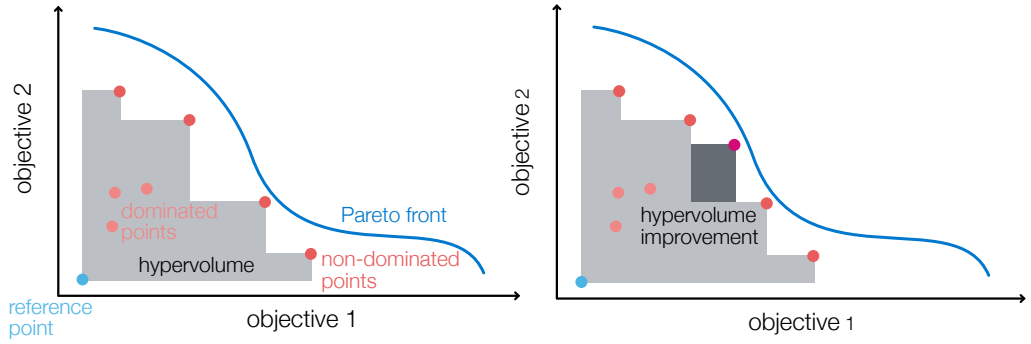

Figure 2.13 – In a multi-objective scenario, the hypervolume—represented here by the light grey area—serves to quantify the quality of the non-dominated solutions, depicted in dark red. When a new non-dominated point is identified, as shown on the right, it contributes to an increase in the hypervolume. Upon convergence, all solutions align along the Pareto front of the objectives, thus providing a comprehensive view of the trade-offs among different optimization goals.

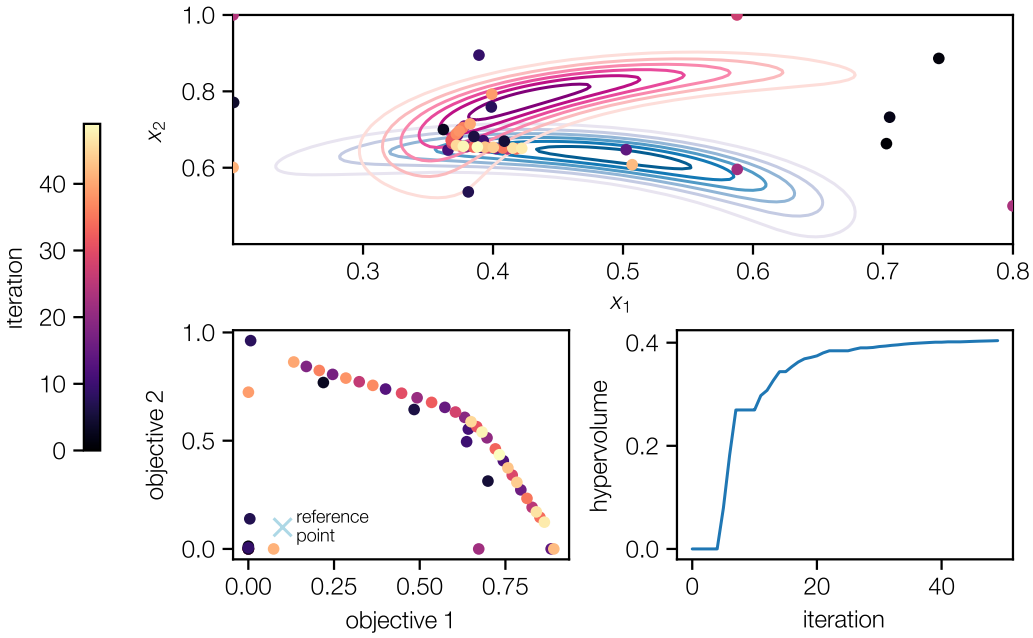

Figure 2.14 – Multi-objective optimization of two competing objectives. Top: input parameter space with unknown objective functions (blue, red, darker is better) and evaluation of the optimization loop (dots) colored by their iteration. Bottom left: Results of evaluations. Bottom right: Improvement of the hypervolume during the optimization process.

Figure 2.14 illustrates an example of multi-objective Bayesian optimization using expected hypervolume improvement on two conflicting objectives. In this instance, the reference point for both objectives is set to 0.1, indicating the desire for both objectives to surpass this threshold. Observing the two objectives, it's evident that the individual maxima of each function are not accepted as solutions because the corresponding other objective falls short in these scenarios.

After an initial exploration, the algorithm identifies the region where both objectives exceed the reference point and begins to sample the Pareto front between them. Given these results, a decision-maker can now choose the desired balance between the two objectives.

2.4 Summary

Bayesian Optimization is a model-based method for finding the optimum of a function that is particularly suited to optimization problems where function evaluations are expensive. The main idea is to use a probabilistic model to predict the objective function and use these predictions to choose where to sample next. It is particularly useful when the function is black-box (we do not have an explicit expression for it), costly to evaluate, or obscured by noise.

At the core of Bayesian optimization is the Gaussian process, which is used to model the function we want to optimize. A Gaussian process is a collection of random variables for which any subset of those random variables has a multivariate normal distribution. It is used to construct a prior distribution over functions, which can be updated into a posterior distribution once data is observed. These updated distributions, reflecting the added knowledge from observed data, are then used to determine the next sample point.

To choose the next sample point, Bayesian optimization uses an acquisition function, which guides where to sample from the space of inputs. This function is derived from the Gaussian process and balances exploration (sampling from regions where the outcome is uncertain) and exploitation (sampling from regions where the model predicts a good outcome). The most commonly used acquisition functions are the Expected Improvement and Upper Confidence Bound. After each function evaluation, the Gaussian process is updated, a new acquisition function is computed, and the process is repeated until the optimum is found.

Multi-objective Bayesian optimization (MOBO) extends the traditional Bayesian optimization concept to handle multiple objectives simultaneously. In real-world scenarios, optimization problems often involve several conflicting objectives. MOBO aims to find the Pareto front, which is a set of optimal solutions where improving one objective will worsen at least one other. Just like single-objective Bayesian optimization, MOBO utilizes Gaussian process models to estimate the objective functions, but it needs to manage a trade-off between multiple objectives rather than a single one.

A crucial part of MOBO is the acquisition function, and the Expected Hypervolume Improvement (EHVI) is a popular choice. The hypervolume measures the volume of the space dominated by the Pareto front. The *improvement* refers to how much an additional sample

could increase this volume. By maximizing the expected increase in the hypervolume, we balance exploration and exploitation across all objectives. Computing EHVI can be challenging due to its complexity, but various efficient algorithms have been proposed to solve this problem.

TERM	DEFINITION
Prior	An a priori assumption about the distribution of a random variable before considering observed data.
Posterior	The updated probability distribution of a random variable after incorporating observed data, calculated using the prior distribution and the likelihood of the data.
Likelihood	A measure of how likely the observed data is, given the parameters of a model.
Marginalization	The process of reducing the dimensionality of a problem by integrating out or summing over certain variables to obtain a marginal probability.
Conditioning	The process of modifying the probability distribution of a variable based on known or observed values of other variables.
Gaussian Process	A stochastic process where any finite collection of random variables has a multivariate Gaussian distribution, often used in machine learning for regression and optimization tasks.
Covariance Function/Kernel	A function that determines the covariance, or the expected level of dependence, between different points in a Gaussian process.
Hyperparameter	Parameters of the prior distribution or the covariance function in a Gaussian process that control the behavior of the model.
Log Marginal Likelihood	The log-likelihood of the observed data, marginalized over the model parameters, used for model selection and hyperparameter optimization.
Surrogate Model	A simplified model that approximates the behavior of a more complex model or a real-world process, used in Bayesian optimization to reduce cost, typically a Gaussian process.
Acquisition Function	A function used in Bayesian optimization that guides the search process by balancing exploration of unvisited regions and exploitation of known good regions.
Hypervolume	In the context of multi-objective optimization, hypervolume is a measure of the space covered by a set of solutions in the objective space, often used as a performance indicator for evaluating the quality of solution sets.

Table 2.1 – Glossary of terms related to Bayesian Optimization and Gaussian Process Regression

3 Bayesian Optimization of Plasma Accelerator Simulations

This chapter will aim to put the concepts outlined in the previous chapters into a combined context. By applying it to particle-in-cell simulations, Bayesian optimization will be used as a tool to unravel the complex interplay of laser and plasma parameters and find optimal configurations for a laser-plasma accelerator setup, like the one employed at LUX.

Section 3.1 will discuss the physical concepts behind the LUX laser-plasma accelerator and the various tuning parameters of the system and their effect on the acceleration process. Section 3.2 will then explain, how with the help of Bayesian optimization, these parameters can be optimized to achieve remarkable beam quality. In section 3.3, the application of multi-objective Bayesian optimization will be discussed.

3.1 LUX laser-plasma accelerator

At the core of every plasma-based accelerator lies a plasma source. This source contains either the gas ionized by the drive laser or a preconditioned plasma generated, for example, by a current discharge [53, 114] or a another laser [115]. Drawing parallels with conventional accelerators, the plasma source simultaneously fulfills the roles of multiple components. The most apparent functions are providing the electric fields to accelerate and guide the electron bunches. As such, the plasma merges the roles of the RF cavity and focusing elements found in conventional accelerators.

Moreover, in most instances, the plasma source also supplies the accelerated electrons, as discussed in section 1.2.3, thereby substituting the gun in a conventional accelerator. This consolidation of functions offers the clear advantage of compactness. However, it may introduce complexity and reduce flexibility, particularly when parameters typically adjusted in separate modules become correlated through shared control parameters.

As a result, to enable precise control over the beam parameters, it is desirable to retain as much task separation and individual tuning capabilities as possible. Concretely, this implies the need to decouple electron injection from acceleration. Typically, this requires spatial and temporal isolation of the injection event. A common method to achieve this, is by inserting obstacles into supersonic gas jets to generate a shock front along which the gas density rapidly decreases [116–118]. The increase of the plasma wavelength corresponding to the change in density causes a local reduction of the wakefield phase velocity and thereby allows isolated injection of electrons (see section 1.2.3). This method, however, relies on the high velocity flow that can be generated with gas nozzles. To handle the load on the vacuum system such jets are typically operated in a pulsed fashion which ultimately limits the achievable repetition

rate and can have implications on the reproducibility [119] and uniformity [120] of the plasma profile.

These limitations can in principle be overcome by replacing the gas jet with a confined volume that is continuously supplied with gas. This can be achieved in capillary or a gas cell based designs. However, generating steep density ramps to trigger injection is not straight forward in these devices which entails the need for different means of generating the accelerated beam. Many early experiments relied on self-injection from wave breaking in very non-linear wakefields [42–44] which can be difficult to precisely control and consequently also lacked good shot-to-shot reproducibility. Currently, a very commonly used injection method, for both gas cells and nozzle targets, is ionization injection. This is partly due to the relatively easy implementation in experiments but also due to the robustness and reliability that can be achieved with it over hours of operation [70, 119].

Also the LUX experiment has been relying on ionization injection since early iterations of its plasma source design. In combination with stringent diagnostic and controls implementations it enabled the first ever daylong operation of a laser plasma accelerator [4].

In its simplest form ionization injection is achieved by mixing a small amount of a high Z dopant gas, typically nitrogen, to the hydrogen or helium that forms the majority of the plasma. Consequently, the dopant gas is available in the entire interaction volume of the laser and the plasma and therefore injection can occur in principle at any time. The continuos trapping of fresh low energy electrons then typically results in large energy spreads. By tailoring the evolution of the driver laser the length over which injection occurs can be reduced but the energy spread that was experimentally achieved this way remains at several percent [121, 47] and is therefore insufficient for the most demanding applications.

Alternatively, when managing to spatially confine the dopant gas within the plasma source the injection event could be isolated and with that smaller energy spreads would be conceivable. Additionally, independent control of beam injection and acceleration could be gained to a certain extent. First realizations of this concept using two distinct gas volumes [122] showed this independent control but were limited to few percent final energy spread.

In the current version of the LUX plasma source the separation of the gas species is achieved through gas dynamics alone so that no distinct second gas volume is required [5, 69]. A similar approach was followed in experiments at BELLA at Berkley lab [123].

In those designs the separation of injection and acceleration in a capillary type design is achieved through the use of multiple supply inlets that provide different gas species. These are either a high Z doped mixed gas for injection or pure low Z gas for acceleration. Typical gases here are hydrogen or helium for the low Z gas and nitrogen for the high Z gas. Depending on the pressure ratio between the inlets, the mixed gas diffuses more or less into the regions that are supplied with the pure gas. When the pressure is roughly balanced a short transition region forms that separates the mixed from the pure gas.

Figure 3.1 shows the cross section of the LUX plasma source design [5] that incorporates this concept. A mixture of hydrogen and nitrogen is supplied through the first (left) inlet and pure hydrogen through the second. This inlet is split before connecting to the axial channel

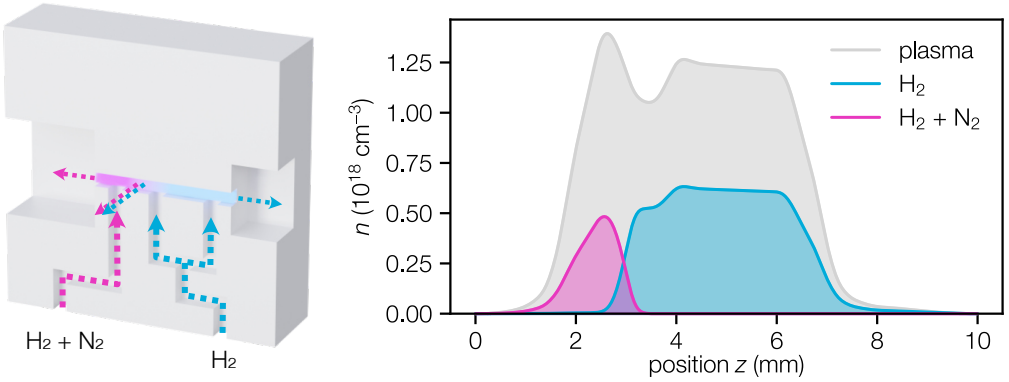

Figure 3.1 – Left: Cross section of the LUX plasma source. The axial channel is supplied with gas via two inlets. The first provides a mixture of hydrogen and nitrogen (magenta). The second, which splits into to inlets supplies pure hydrogen (blue) to form a plateau in the back of the plasma source. A gas outlet near the middle of the channel prevents mixing of the two gas species. Right: Corresponding density profiles when using 10% nitrogen in the mixed gas.

to provide a wide plateau of approximately constant density. A hole on the side of the main channel (not shown in the figure) acts as gas outlet that further prevents mixing of the two species in the plasma source.

Multi-species computational fluid dynamics (CFD) simulations that were used during the design of the setup (see [69] for extensive discussion) confirm that this concept in fact leads to a strict separation of the mixed and the pure gas when correctly setting the pressures of the two inlets. The plot in fig. 3.1 shows the axial density profiles of the mixed and pure gas and the resulting plasma density when considering ionization of the hydrogen and the outer shells of the nitrogen.

In the fluid simulation the first inlet was supplied with a 9:1 mixture of hydrogen and nitrogen. The second inlet was supplied with pure hydrogen at a slightly higher pressure than the mixed gas. As shown in fig. 3.1, this configuration leads to a clear separation of the mixed from the pure gas with a short transition region that is roughly 1 mm long. The pure gas profile shows a slight drop at the transition point which is caused by the gas outlet at this point. The resulting plasma profile has a peak at the beginning which is caused by the five outer shell electrons of the nitrogen atoms that are ionized together with the hydrogen by the leading edge of the drive laser. Behind the peak the gas outlet creates a lower density valley that creates a short upward ramp into a flat plateau which has a slight downward density gradient.

3.1.1 Simulations

To further elaborate and study laser-plasma acceleration in gas profiles like the one shown in fig. 3.1, particle-in-cell (PIC) simulations using the PIC code FBPIC [124] are performed. FBPIC is specialized for simulating plasma based accelerators through two distinct features.

First, it exploits the fact that many LPA related problems have an approximately cylindrical symmetry, i.e., the wakefield is roughly rotation invariant. However, instead of modeling just purely cylindrical fields, FBPIC models an azimuthal Fourier decomposition of the fields [125]. This allows to simulate also fields that have azimuthal dependencies like the driver laser that is typically linearly polarized. If the azimuthal dependency is not too complex, as in the case of the laser, the Fourier decomposition can be truncated after a couple of modes. With this a simulation that captures the relevant three-dimensional effects can be simulated at the cost of a few 2D simulations.

The second important feature is that it is possible to run simulations in a Lorentz boosted frame which allows orders of magnitude speedups [126] by reducing the strong discrepancy of lengthscales that need to be resolved when modeling for example the small laser wavelength during the propagation through the millimeter long plasma cell. The application of the boosted frame is enabled through the use of a spectral solver and a special Galilean transformation [15, 127, 10] that prevent the deteriorative effects of numerical Cherenkov radiation [13] and the growth of the numerical Cherenkov instability.

If not stated differently, the simulations presented in the following use a grid resolution of $\Delta_z^{-1} = 50\,\mu\text{m}^{-1}$ and $\Delta_r^{-1} = 3\,\mu\text{m}^{-1}$ and a time step of $\Delta_t = \Delta_z/c$. Two azimuthal modes are used to capture the cylindrical symmetry of the wakefields and the linearly polarized drive laser. The simulations are performed in a Lorentz boosted frame with $\gamma = 5$. The laser beam in the simulations has a Gaussian temporal profile and the radial mode of a real high power laser is approximated as a flattened Gaussian beam (order N=100) [128] which is defined bei a weighted sum of cylindrically symmetric Laguerre-Gauss modes.

The hydrogen atoms in the simulation are modeled with 8 macro-particles per cell. To improve the statistics on the ionization-injection process the nitrogen is represented by 16 macro-particles per cell. The hydrogen and the first 5 electrons of the nitrogen, i.e., its outer shells, are assumed to be ionized by the leading edge of the drive laser and are therefore pre-ionized in the simulation. The field ionization of the inner nitrogen shells is modeled according to the Ammosov-Delone-Krainov (ADK) model [77, 78].

3.1.2 Localized ionization-injection

To discuss the dynamics and unique properties of the plasma source concept in section 3.1, a PIC simulation with laser parameters like they are typically used at the LUX experiment is performed. Accordingly, a drive laser with 2.65 J with 34 fs FWHM pulse duration is focused 5.41 mm behind the entrance of the plasma source to a vacuum spot size of 25 µm.

The evolution of the wakefield and the longitudinal phase space (E, ξ) is shown in fig. 3.2. When the laser, which still has a large radius and consequently low intensity, enters the plasma (1), it drives a nearly linear sinusoidal wakefield. Despite the gas being already doped with nitrogen atoms, the low amplitude of the wakefield and more importantly, the positive density gradient of the plasma, inhibit the injection of electrons into the wakefield. As the density of the mixed gas species peaks, the plasma density gradient inverts. This shift induces a change

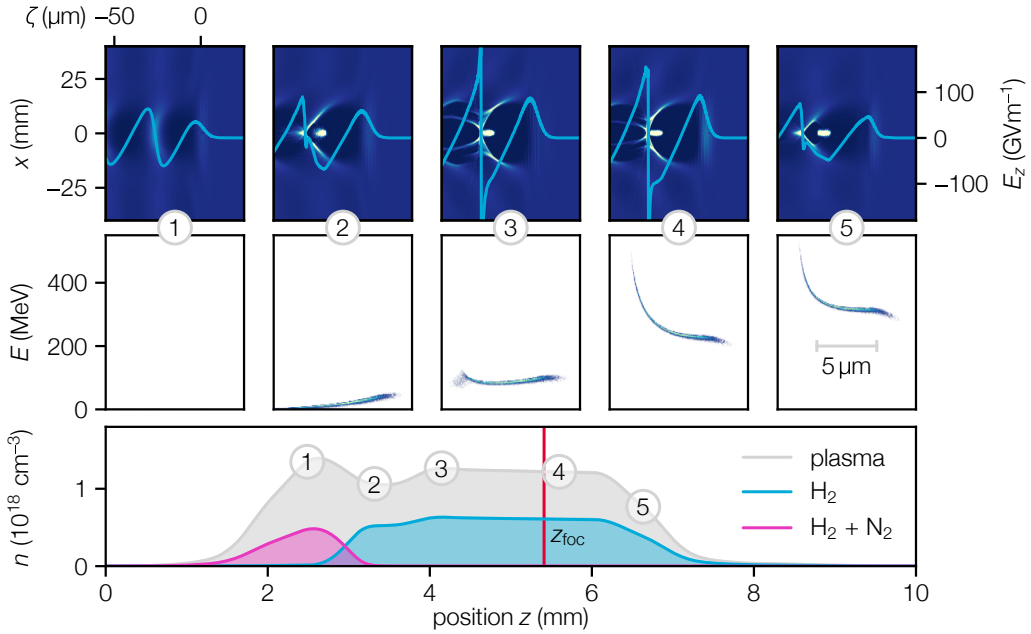

Figure 3.2 – Evolution of the wakefield and longitudinal phase space in a plasma source with localized ionization-injection. Upper row: Charge density with on-axis accelerating field. Middle row: Longitudinal phase space (E, ξ) for different positions in the plasma. Bottom plot: Plasma and gas densities with markers for the positions of the corresponding plots above. The vacuum focus position of the drive laser z_{foc} is marked by the red line.

in the plasma wavelength, reducing the phase velocity of the wakefield, which then initiates electron injection. As a result, an electron bunch starts to form.

Due to the ongoing elongation of the plasma cavity in the density downramp, the phase at which new electrons are trapped continues to advance. Therefore, electrons injected at different times occupy different phases of the wakefield. Consequently, a lengthy electron bunch with a small slice energy spread is formed at the end of the downramp (2), where the transition between the mixed and pure gas occurs. The bunch exhibits a positive correlation (or chirp) between the longitudinal bunch coordinate and the energy. This correlation is primarily a result of the shifting phase during injection. Electrons in the head of the bunch are injected first, and thus, experience more acceleration, resulting in higher energy.

Following the gas transition region is a slight density upramp (3) wherein the wakefield contracts a bit, causing some electrons at the bunch's end to be defocused and decelerated. The tail of the bunch is now positioned at the very rear of the plasma cavity where the accelerating field is the strongest.

At this point, as the laser nears the focus, its intensity significantly increases, driving a strong, non-linear wakefield in the blowout regime. Although the beam charge influences the longitudinal electric field through beam loading, the field is not entirely flattened, leading to the tail of the bunch receiving more acceleration than the head. As a result, the longitudinal phase space

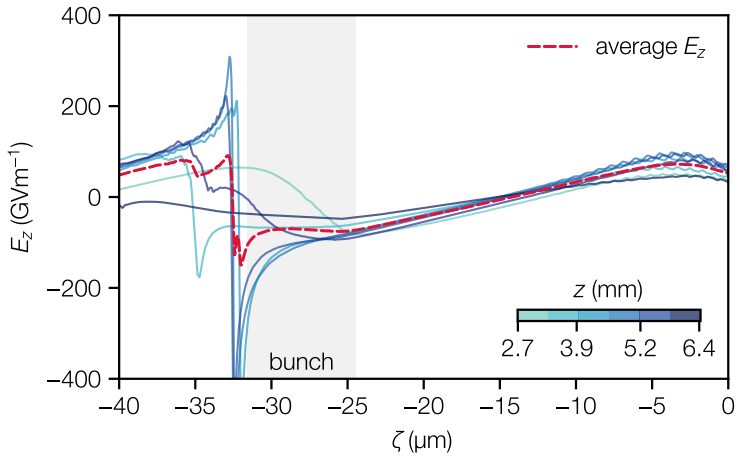

Figure 3.3 – Evolution of the accelerating field during the propagation through the plasma. Variations of the beam loading effect due to changes of the wakefield strength result in on average flattening of the field (red dashed line) over the majority of the bunch length (grey area). The colors of the lines indicates the corresponding position in the plasma source, see colorbar.

undergoes a shearing motion and its correlation decreases. As the plasma density plateau concludes (4), the energy correlation is fully compensated and even slightly inverted.

However, soon after, at the plasma profile exit (5), the laser diffracts and the wakefield weakens. Now, due to overly strong beam loading, the bunch reverses the slope of the accelerating field so that the tail is accelerated less than the head. This results in the flattening of the majority of the longitudinal phase space. The bunch has a tail with higher energies, which is attributed to the field spike at the end of the plasma cavity. These electrons contribute less than 10 % of the 66 pC of total charge.

The longitudinal field profiles in fig. 3.2 show that the setup does not operate at optimal beam loading conditions, i.e., flat accelerating field along the bunch, at every given point in time. This is not possible since the beam current is constant after the injection while the wakefield amplitude changes due to the evolution of the drive laser. However, as indicated above, what is relevant for the final phase space of the bunch is the average beam loading field. This means that non-ideal beam loading at one position is acceptable as long as it is compensated somewhere else.

This concept is illustrated in fig. 3.3 which shows the longitudinal field at a few positions in the plasma and the average field between the point of injection and the end of the plasma. The combination of over- and underloading at different positions leads to a field that is flat over the majority of the bunch on average. Additionally, a perfectly flat field would not even be desirable in this case, because the energy correlation that the bunch has after injection needs to be compensated. Therefore, the ideal field has a slight slope that corrects this correlation over the length of the plasma.

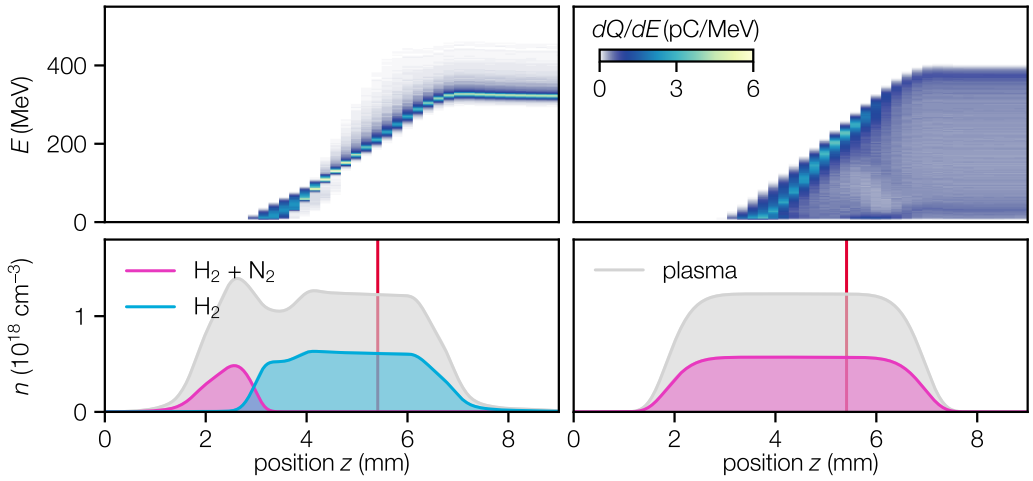

Figure 3.4 – Evolution of beam energy spectrum in a plasma with and without localization of the dopant gas. To get comparable beam charges the setups provide the same total amount of nitrogen over the full gas profile. Consequently, the nitrogen concentration in the mixed gas is 10 % in the localized and 1.96 % in the non-localized case. The laser focus position is marked by the red line.

The benefits of localized over non-localized ionization-injection in terms of beam quality become evident when comparing the evolution of the electron beam energy spectrum in both cases. Figure 3.4 shows the energy spectra from the simulation above for each z Position along the plasma and compares them to a simulation where the nitrogen is homogeneously distributed over the full length of the gas profile. To provide the same amount of injectable charge the total amount of nitrogen is the same in both simulations. Consequently, the nitrogen ratio in the mixed gas was reduced from 10 % to 1.96 %.

In the case with localized mixed gas the injection event is limited to the presence of nitrogen and is further constrained to the region where the plasma density is decreasing. Consequently, since no further electrons are injected afterwards a single energy isolated bunch is formed that is accelerated to higher energies in the density plateau.

In the scenario of continuous nitrogen doping, the initially injected beam loads the wakefield, which subsequently diminishes the injection rate. Despite this, injection does not halt completely, and lower-energy electrons continue to be injected. This results in a spectrum that peaks at higher energies due to the early-injected electrons, along with a tail extending across the entire lower energy range. Near the end of the plasma profile, the wakefield becomes overloaded with charge, causing some early-injected electrons to decelerate. Consequently, the high-energy portion of the beam loses charge, leading to a decrease in spectral density. The outcome is a wide spectrum with the charge evenly distributed over all energy levels.

3.1.3 Tuning parameters

A distinct advantage of dividing responsibilities within the plasma source, such as injection and acceleration, is the possibility to individually fine-tune these critical elements. However, it's important to bear in mind that these elements aren't entirely independent from one another. For example, the injection process can influence the acceleration of the beam through effects like beam loading. Yet, being equipped with the correct distinct control parameters means that in principle both aspects can be adjusted to each other to give the best end result. The following paragraphs will discuss the most important parameters of the setup and their effect on the accelerated electron beams.

As the acceleration of electrons in an LPA is the result of the interaction between a laser and a plasma naturally the parameters of both these constituents play important roles. The drive laser is mainly characterized by its energy as well as its temporal and spatial profiles. Although, higher order aberrations of the spatial and temporal phase of the beam can have an impact on the accelerated beams [5, 129], the most relevant properties in this context are the pulse duration τ_{laser} and focus position z_{foc}. Additionally, the focal spot size of the beam is important. However, since the focal length is fixed in experiments there is typically not much dynamic control over it. For this reason the laser spot size is not considered a tuning parameter in the following despite its importance in the designing phase of the accelerator.

The gas profile as well as the gas composition in the plasma source is controlled via the flow and mixing ratio of the gases going into the inlets. This influences the overall density of the resulting on-axis plasma and the amount of nitrogen in the injection region. Further, the ratio of flows between the first and second inlet determines the position and the length of the transition region between mixed and pure gas. For simulations, the complex gas dynamics involved when changing the setpoint of the plasma source would ideally warrant a full multi-species fluid dynamics (CFD) simulation to determine the resulting gas profile. Unfortunately, the setup of these simulations is currently still a process that involves some manual tuning. Implementation of this into fully automated start to end simulations including the CFD part is currently being worked on, though. As an approximation, the CFD simulation for the gas profile that is shown in fig. 3.1 is scaled and used for the PIC simulations. In particular, the nitrogen concentration in the mixed gas and the overall gas density is varied. It was confirmed through CFD simulations that the resulting gas profiles are feasible. Nevertheless, it should be noted that this does not represent all possible degrees of freedom of the plasma source.

Figure 3.5 shows the effect of the most important laser and gas parameters on the longitudinal phase space (E, ξ) of the accelerated electron bunch. Changes of the different parameters have very distinct effects on the properties of the bunch during injection, acceleration or both. These effects will be explained in the following going row by row in fig. 3.5.

Laser energy—Altering the laser energy influences both the injection and acceleration stages of laser-plasma interaction. Initially, higher energy, and consequently higher intensity in the mixed-gas region of the plasma source, leads to increased ionization of nitrogen atoms. In addition, the increased wakefield strength enlarges the injection volume, which is the area

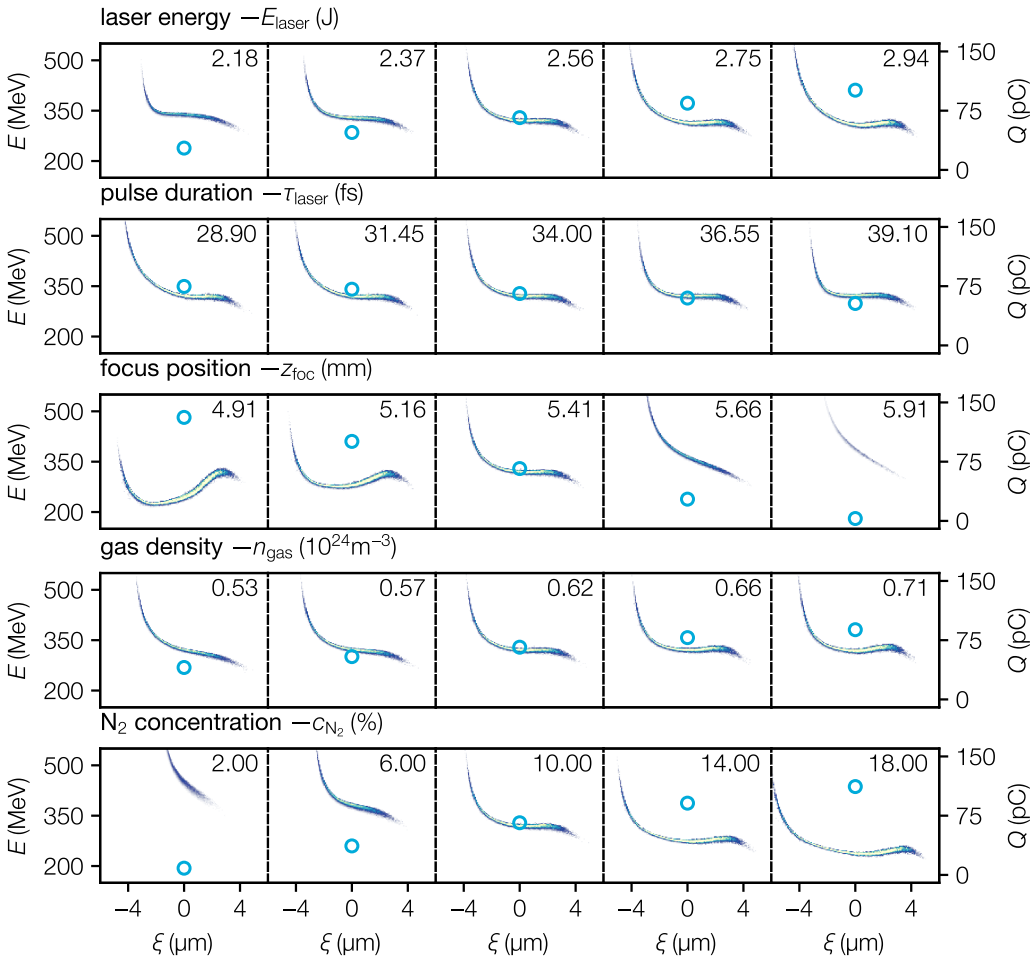

Figure 3.5 – Effects of laser and gas tuning parameters on the longitudinal phase space of the accelerated electron bunch. The center column corresponds to the setpoint illustrated in fig. 3.2. The blue dots indicate the charge of the bunch.

where nitrogen atoms are ionized and the wakefield potential is sufficiently high to capture the released electrons. Consequently, the beam charge increases with laser energy.

As the wakefield strength in the acceleration section also intensifies, the beam loading remains roughly matched, meaning the linear part of the energy correlation in the bunch does not significantly shift. However, observable changes occur in the higher-order correlations in the (E, ξ) phase space, imparting a slightly parabolic curvature to the beam when the driver energy is increased.

Pulse duration—Similar effects are present when tuning the pulse duration of the driver, since this as well is a parameter that foremost determines the intensity of the pulse. However, in the range shown here, ±15% for both parameters, the effects of changing the pulse duration are less pronounced. This is explainable by the fact that the leading edge of the drive laser is

more or less steep depending on the pulse duration. With this the phase of the wakefield also slightly changes. This slight shift of the wakefield with respect to the driver partly counteracts the increased ionization volume.

Focus position—The longitudinal focus position determines the size and intensity of the drive laser in the injection region. As the focus is located downstream of the mixed gas region, the laser spot size w_0 at the point of injection is approximately a linear function of the focus position. This implies that the ionization volume also changes linearly with the focus. As a result, shifting the focus downstream or upstream by a few hundred micrometers can significantly vary the beam charge.

In the acceleration region where the focus is located, shifting the focus position on this scale has a negligible effect. This is primarily because effects occurring in front of the focus will cancel out behind it, leaving the average wake amplitude unaffected.

In combination, altering the charge amount without affecting the wakefield results in different amounts of beam loading. As shown in fig. 3.5, this leads to a linear chirp of the phase space when deviating from the optimum. Due to its minor impact in the acceleration region of the setup, the focus position can be used to match the beam charge to optimize beam loading in a given wakefield.

Gas density—The overall gas density dictates both the plasma density and the quantity of nitrogen atoms that can be ionized, which in turn corresponds to the charge that can be injected. The plasma density directly scales the accelerating field and impacts the self-focusing of the drive laser. Taken together, an increase in plasma density results in the enhanced acceleration of the bunch.

As shown in fig. 3.5, there's a slight increase in energy at the head of the bunch with rising gas density. The tail of the bunch doesn't experience this increase in accelerating fields due to the stronger beam loading from the additional charge that's injected.

Nitrogen concentration—The final free parameter under consideration is the dopant gas concentration. At first glance, one might assume that the concentration of nitrogen in the mixed gas would primarily scale the charge of the bunch. However, given the five outer shell electrons of nitrogen that are already ionized in the leading edge of the driver, the dopant concentration becomes essential for shaping the resulting plasma profile.

Specifically, the peak at the start of the plasma profile is primarily due to electrons from nitrogen atoms. Hence, an increase in nitrogen concentration leads to a larger density difference in the plasma ramp that initiates the injection. This brings about a more substantial change in the plasma wavelength, resulting in the beam being injected over more phases of the wakefield, ultimately leading to a longer bunch length.

The rear of the wakefield at the start of the injection roughly determines the position of the head of the bunch. With a higher initial plasma density, the bunch head is positioned closer to the drive laser. As a result, in the second half of the plasma source, it will be situated in a wakefield phase with a lower accelerating field, leading to a reduced final energy.

3.2 Optimization

The intricate relationship between tuning parameters and beam parameters, as detailed in the preceding section, is a feature of nearly all LPA setups. This complexity presents a significant challenge when it comes to identifying an optimal set of input parameters that deliver superior beam quality for a given task. For instance, demanding applications such as free-electron lasers may require minimizing the energy spread of electrons while maximizing the beam charge or current. However, other potential uses of LPAs may necessitate maximizing the beam energy or the energy transfer efficiency between the laser and the accelerated electrons.

Moreover, the optimal set of input parameters could change due to iterative modifications of the setup, such as adjustments to the plasma source geometry, or due to day-to-day variations in uncontrolled environmental parameters in experiments. Consequently, a single setup might require optimization on multiple occasions. Given the demanding and costly nature of probing a setpoint in an LPA, either through Particle-In-Cell (PIC) simulations or experiments, it becomes imperative to employ efficient numerical optimization methods. Therefore, as discussed in chapter 2, Bayesian optimization is an obvious approach to this problem.

In the following the application of Bayesian optimization to produce high quality electron beams will be demonstrated using PIC simulation of the LUX LPA setup that was discussed in section 3.1.

3.2.1 Objective function

To set up an optimization workflow, the first step is to define an objective, which is the optimization goal. This function maps the outcome of an optimization step to a scalar quantity that is then optimized. In this case, the physical goal is to achieve optimal beam loading conditions, which involves flattening out the average accelerating field while transporting maximum beam charge. A reasonable objective is a combination of beam charge q and relative energy spread $\Delta E/E$. However, since it is often easier to increase beam charge than to decrease energy spread, a simple objective of $f(\boldsymbol{x}) = q/(\Delta E/E)$ would tend to find beams with high charge at the cost of non-ideal beam loading. Conversely, if only energy spread is used as the objective, the optimizer would find settings resulting in zero charge. To prevent this, the objective function

$$f(\boldsymbol{x}) = \sqrt{q[\mathrm{pC}]}/(\Delta E/\tilde{E})[\%] \tag{3.1}$$

is used, where the charge is scaled with a square root to encourage the optimizer to find beams with small energy spread. The goal of the optimizer is to maximize this objective function.

It's important to acknowledge that this objective function was selected primarily through domain knowledge and preliminary trial-and-error iterations. Even though the function is grounded in physical reasoning, it's not definitive in the sense that it holds ambiguities regarding the final beam parameters. A more systematic approach, entailing the optimization of multiple beam parameters using multi-objective optimization, will be explored later in this chapter.

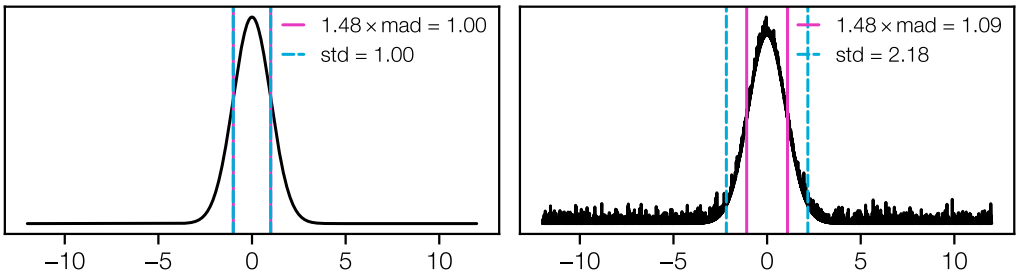

Figure 3.6 – Comparison of the standard deviation (std) and the median absolute deviation (mad) for ideal (left) and obscured signals (right). For comparison, the mad is scaled by 1.48 to match the std in the ideal case.

For efficient optimization it is important to define an objective function that is robust against noise and chaotic outliers. Otherwise, the surrogate model of the Bayesian optimizer needs to introduce a larger noise level to interpret the data (see section 2.2) which can lead to overly exploited regions of the parameter space. A common definition for the energy bandwidth of an electron bunch is the standard deviation (std)[1]. However, this definition is prone to overestimation of the energy spread in the presence of noise or small tails in the spectrum. The *median absolute deviation* (mad) is a related width measure defined as

$$\text{mad} = \text{median}(|X_i - \tilde{X}|) \tag{3.2}$$

where $\tilde{X}$ is the median of the dataset $X_1, \ldots, X_n$. Due to its linear dependence on the distance compared to the quadratic one of the standard deviation the mad is more robust against outliers that are far away from the main signal.

The robustness of the median absolute deviation is demonstrated in fig. 3.6 where for comparison the mad is scaled to match the standard deviation in the ideal case of a gaussian signal. Once noise is added to the signal the standard deviation more than doubles while the mad increases only slightly. Naturally, this is a constructed example, and improvements could easily be made by limiting the range over which the signal is analyzed or employing denoising methods.

However, during the optimization process, the shape and position of the signal, i.e., the energy spectrum of the electrons, can substantially vary based on the input parameters. This variation makes it challenging to anticipate suitable post-processing methods. Consequently, to circumvent this issue, robust measures for the energy bandwidth (ΔE) and central energy ($\tilde{E}$) are employed. As defined in the objective eq. (3.1), the objective uses the mad and the median, respectively, to evaluate ΔE and $\tilde{E}$. This choice of robust statistics helps ensure the performance measure isn't significantly swayed by potential outliers or noise in the data.

[1] The standard deviation of the energy spectrum is commonly also referred to as the rms energy spread.

Table 3.1 – Search space used for the Bayesian optimization of the LUX inspired LPA setup.

laser energy	E_{laser}	$1.8 - 2.7$	J
laser pulse duration	τ_{laser}	$34 - 51$	fs
focus position	z_{foc}	$3 - 7.5$	mm
gas density	n_{gas}	$0.37 - 0.74$	10^{24}m^{-3}
N_2 concentration	c_{N_2}	$1 - 15$	%

3.2.2 Search space

The optimization process also requires defining the search space that the optimizer should explore. Although Bayesian optimization is known for its sample efficiency, it can still fall victim to the phenomenon known as the *curse of dimensionality*. This term refers to the fact that the volume of the search space grows exponentially with the number of input parameters, dramatically increasing the cost, i.e., the time required to find the optimum. Consequently, a balance needs to be struck between using a few parameters to get a result in a reasonable time, while still ensuring enough degrees of freedom to discover an optimal configuration.

Table 3.1 presents the parameters and their respective ranges that define the search space employed in the following sections. These parameters are the same ones discussed in section 3.1.3. The parameter ranges were chosen based on estimates where acceptable beam quality could be expected and what is achievable in the experiment. These ranges serve as constraints in the optimization process, guiding the optimizer towards feasible solutions.

The importance of using efficient methods like Bayesian optimization becomes evident when estimating the number of simulations that would be required when trying to find the global optimum using a systematic grid scan. Assuming that in order to find the global optimum all parameters have to be tuned to a precision of 5 % of the full search space, the number of configurations to test would be $20^5 = 3.2{\times}10^6$. With a runtime of ∼30 min per PIC simulation[2] such a systematic scan would consume 1.6×10^6 h of computing time.

3.2.3 Bayesian optimization of localized ionization injection

Upon establishing the setup, search space, and an objective function, the problem can be optimized. This is accomplished using Bayesian optimization with the *q-expected-improvement* acquisition function [105], as implemented in the BOTORCH framework [103]. When optimizing with PIC simulations, the advantage of running multiple evaluations in parallel can be capitalized upon. The q in the name of the acquisition function represents its capacity to evaluate using a batch of q concurrent simulations. However, the degree of parallelism that makes sense is limited, as Bayesian optimization derives its efficiency from the information used to construct the surrogate model. If many evaluations are run in parallel, the optimizer

[2] Run on a single nvidia A100 GPU on the DESY Maxwell cluster

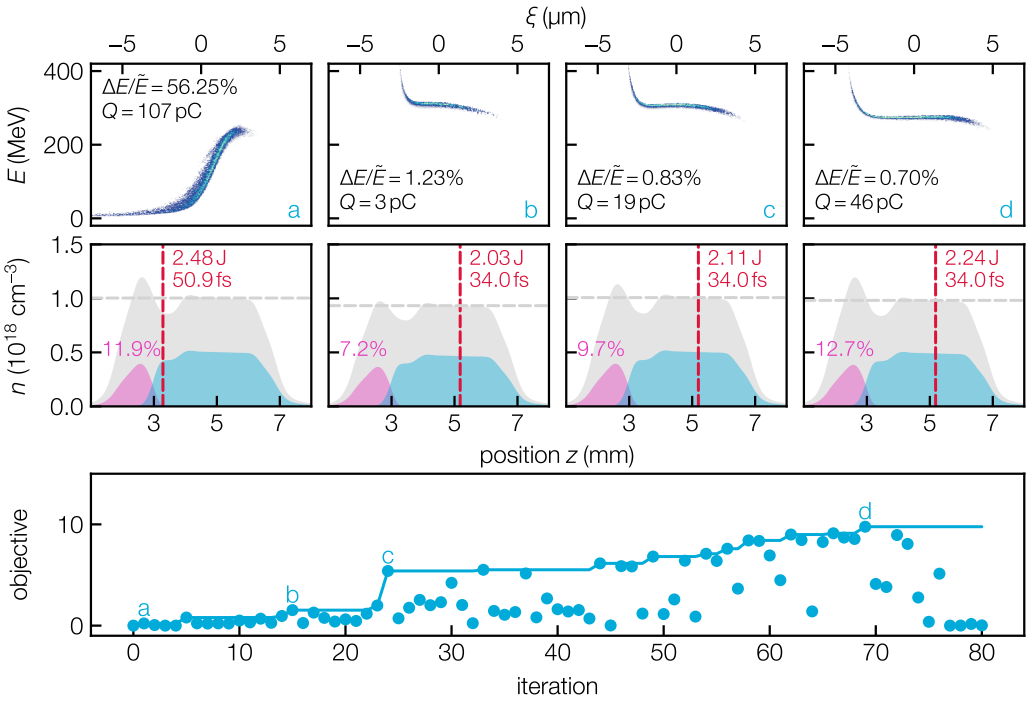

Figure 3.7 – Optimization of the beam quality of electrons accelerated in the LUX plasma source. Bottom: Objective function values (eq. (3.1)) of individual simulations during the optimization (dots) and cumulative best result (solid line). Middle row: Selected setups tested by the optimizer. Shown are the profiles of the mixed gas (magenta) with nitrogen concentration, the pure hydrogen (blue) and the resulting plasma density (grey) as well as the position of the laser (dashed line) with the pulse energy and duration. Top row: Resulting longitudinal phase spaces ($\xi - E$). The corresponding iterations are indicated by the letters in the objective plot. The optimization was run for 400 iterations but only the first 80 are shown here after which no further improvement was observed indicating that the global optimum was found.

suggests many setpoints based on the same information, thereby reducing the utility of each of these suggestions. In this study, $q = 4$ simulations are executed concurrently, offering an decent trade-off between computational efficiency and information gain.

The results of 80 iterations of Bayesian optimization are presented in fig. 3.7. To begin the process, six randomly selected configurations are simulated, which are then used to generate a rough Gaussian process surrogate model of the problem.

The beam resulting from these initial evaluations is shown in fig. 3.7(a). It exhibits a very broad energy spectrum, ranging from 0 MeV to nearly 300 MeV, while also having a charge of 107 pC. The high charge suggests that the broad spectrum is due to overly strong beam loading, where the wakefield is reversed, resulting in less acceleration or even deceleration of the bunch tail. The suboptimal conditions of this configuration stem from the laser focus being too close to the mixed gas region. The high laser intensity leads to a large ionization volume

and the injection of a large amount of charge that is not supported by the weak wakefield, which is due to the long pulse duration of 51 fs.

Following the initial random steps, the optimizer selects the subsequent input parameters using to the acquisition function, i.e., points with a high expected improvement over the current optimum according to the surrogate model. With this, after 15 evaluations a first configuration is found that gives a relatively flat phase space (fig. 3.7b). The excessive beam loading caused by too much charge is compensated by reducing the laser energy and shifting the focus downstream and thereby reducing the intensity in the mixed gas region. Additionally, the nitrogen concentration is reduced and with that the amount of injectable charge. In combination, the bunch charge is reduced to only 3 pC, which however matches the wakefield for this configuration and gives a phase space that is flat over a large fraction of the beam. Compared to the phase space in fig. 3.7a the beam is much shorter. This is caused by the lower nitrogen concentration and the resulting low plasma density in the injection region of the plasma. The peak in the beginning of the plasma now has almost the same plasma density as the following plateau. The plasma wavelength at this position where injection starts to occur determines the phase of the head of the bunch and the wavelength in the plateau the phase of the tail, i.e., the bunch is truncated by the defocusing region of the wakefield. The fact that the plasma densities and with that the plasma wavelengths are almost balanced in this configuration results in the short bunch length.

After ten more evaluations, the optimizer finds a configuration that gives a significant improvement of the objective (fig. 3.7c). This is the result of again increasing the N_2 concentration and slightly raising the laser energy. As a consequence, the charge increases to 19 pC and the beam becomes longer again. With this, the longitudinal phase space is now flat over a larger fraction than before which reduces the projected energy spread to 0.8 %.

Over the course of the next fifty iterations the objective is gradually increased before reaching a maximum after a total of 70 evaluations (fig. 3.7d). To get there, the nitrogen concentration is further increased while the beam loading is kept at an optimum by increasing the laser energy to account for the additional charge.

Figure 3.8 shows the phase space resulting from this optimum in more detail. The increase of the nitrogen concentration leads to a long beam with 8 μm head to tail bunch length which combined with a charge of 46 pC gives a peak current of 2.5 kA. By optimizing the wakefield to match the beam loading generated by the bunch, the majority of the longitudinal phase space is flattened resulting in a projected energy spread of 0.7 % (mad). The standard deviation of the spectrum is comparably large with 4.5 % which is caused by the curled up tail of the bunch. This tail accounts for less than 10 % of the beam charge and is a result of the strongly non-linear field near the back of the wakefield bucket. Figure 3.8 also shows the transverse phase spaces of the optimal setup which were not directly optimized. As expected from ionization injection, the horizontal emittance, i.e., the polarization plane of the laser, is larger than the vertical.

The observed beam emittance, despite not being a direct target of the optimization, remains relatively small with values of 1.4 and 0.3 mm mrad. This could be attributed to the optimization's focus on minimizing the energy spread. Both energy spread and beam emittance are affected by the injection volume [80], suggesting a connection. If the beam emittance is of

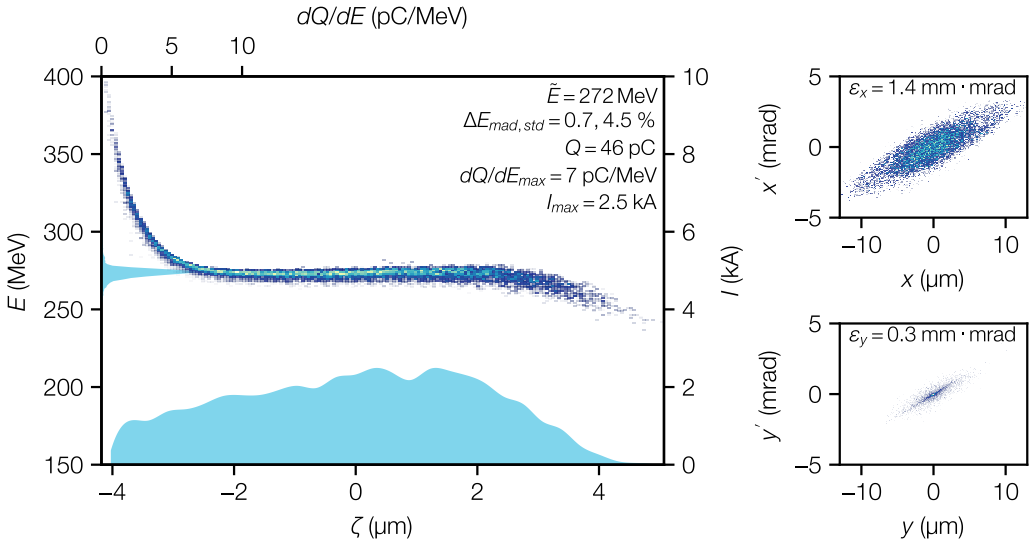

Figure 3.8 – Longitudinal (left) and transverse (right) phase spaces from the optimized setup. The current profile and the energy spectrum are shown as projections on the x and y axis, respectively.

critical importance for the intended application, it can be included in the objective function for direct optimization. This option will be explored further later in this chapter. Indeed, the task of experimentally measuring emittance is relatively more intricate compared to assessing charge and spectrum, leading to a higher level of complexity when it's incorporated into optimization procedures.

Figure 3.7 illustrates a rapid decrease in the objective values following the discovery of the optimum. This pattern suggests the optimizer is shifting toward a more explorative approach, having identified that the region surrounding the perceived optimum no longer offers significant anticipated improvement. The optimization process was carried out over 400 iterations without discovering a better configuration, suggesting that the identified optimum is likely a global optimum within the search space, or at least in close proximity to it.

3.2.4 Surrogate model based stability analysis

The subtle adaptations of the tuning parameters required to reach the improvement from fig. 3.7b to fig. 3.7d indicate a certain sensitivity of the system to at least some of the inputs. In practice, this can have severe implications on the achievable stability because of unavoidable fluctuations of the input parameters. Therefore, proper analysis of the behavior of the objective and its response to input variations is highly desirable. Typically, this would require to run numerous additional simulations to scan the surroundings of the optimum or more ideally even simulate the expected jitter with sufficient statistics. This task can quickly add up to hundreds or thousands of additional costly evaluations.

Table 3.2 – Lengthscale hyperparameter of the BO surrogate model. Values are given in SI units and normalized to the search space range to show the relative importance of each parameter.

		SI	NORMALIZED
laser energy	E_{laser}	0.15 J	0.16
laser pulse duration	τ_{laser}	17.5 fs	1.03
focus position	z_{foc}	0.05 mm	0.01
gas density	n_{gas}	$0.02\,10^{24}\text{m}^{-3}$	0.07
N_2 concentration	c_{N_2}	2.05 %	0.15

For a more efficient way of estimating the behavior of the system, an extremely valuable property of Bayesian optimization is that it provides insight beyond the optimization result through its model based nature. The surrogate model that is used to guide the optimization also allows to make statements about the importance of the individual tuning parameters. Further, it allows to make quick predictions about the system at points that have not yet been evaluated. With this the robustness of the optimization result to variations of the input parameters can be estimated.

One benefit that Gaussian process models have over other complex regression methods like neural networks is that their hyperparameters are actually describing the characteristics of the data. Recall from section 2.2.2 the definition of the RBF kernel:

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x}'|^2}{2l^2}\right). \tag{3.3}$$

Actually the surrogate model for the optimization in section 3.2.3 uses a Matérn kernel but the reasoning is the same. The kernel of a Gaussian process model describes the similarity between the points $\boldsymbol{x}$ and $\boldsymbol{x}'$. The lengthscale hyperparameter l describes over which length this similarity decays. Consequently, if l is large the modeled function will only slowly change as the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$ increases. On the contrary if l is very small it may change rapidly. For a multi dimensional model like the one used for Bayesian optimization l is a vector with one entry per tuning parameter. It therefore describes the sensitivity of the objective to variations in each of the input parameters.

The lengthscale parameters from the surrogate model of the optimization above after 400 iterations are shown in table 3.2. The values for l are shown in the corresponding units of the input parameters and normalized over the range of the search space. The former can be seen as a rough lower estimate on how much precision and stability of the tuning parameters is required to hit the found optimum reliably. The normalized representation gives an impression about the importance of each parameter for the optimization over the given search space. For instance, the normalized lengthscale for the pulse duration is close to one, indicating that within the range of the optimization this parameter has only very little influence. According

to this analysis the focus position is by far the most influential parameter, where the objective changes significantly for variation on the order of 1 % of the search space.

A more detailed impression of the interplay between the tuning parameters and the beam quality can be gained by evaluating the surrogate model directly. Figure 3.9 shows the predicted objective function by the model for slices through the search space around the optimum. The first thing that stands out is again the high sensitivity of the objective to changes of the laser focus position. This is probably best seen in the scan over z_{foc} that is shown in the center of the figure. The region over which the objective is large is extremely narrow compared to the range over which the setup was optimized. In fact, the objective falls of to 50 % over just 100 μm of focus shift.

Apart from the relatively simple information about the importance of the tuning parameters fig. 3.9 indicates some more complex interplay between the variables. For example, the scan over the nitrogen concentration and the laser energy (fourth row, first column) shows an interesting behavior. Starting from the optimum (X), when reducing the N_2 concentration, the objective decreases as one would expect from the loss of charge associated with it. However, at the same time the dependency on the laser energy seems to become less strong as indicated by the widening of the bright patch along the x-axis. This hints towards the fact that with this setup tradeoffs between the beam quality and the stability of the system can be made by adapting the setpoints. A deeper discussion on this aspect will follow later in this chapter.

Figure 3.9 also again gives an impression about the complexity of the optimization problem. Although only showing a very small fraction of the total search space volume, the individual slices already show large dark regions, i.e., regions where the objective is expected to be small. Consequently, beams with high beam quality can be expected only in a very small volume of the parameters space.

The predictive power of the surrogate model should however be treated with caution. This is because the data that the surrogate model is conditioned on was taken with the purpose of finding the optimum and not maximising the information about the system. For example, the optimizer would not try to sample regions in which the objective is expected to be very small as these values would probably give low values of the acquisition function. Consequently, this means that the surrogate model is likely inaccurate in these regions. The beauty of using a probabilistic model, however, is that these inaccuracies or uncertainties are quantified and with that the level of trust that can be put into the model.

Making predictions far away from the training data is most inaccurate when the lengthscale of the model is particularly small as for the focus position in this example. In this case the correlation between two points is only present over a small distance. Consequently, a sampled data point provides predictive information only for a very confined region around it. For the focus position it can be seen in fig. 3.9 that when moving away from the sampled point in the center of the peak the uncertainty quickly increases as one moves away from it. Additionally, in $(z_{\mathrm{foc}}, n_{\mathrm{gas}})$-plot discrete islands of high objective values can be seen. Although this is physically thinkable it is more likely a result of the high degree of localization of information and missing data to model the correct behavior.

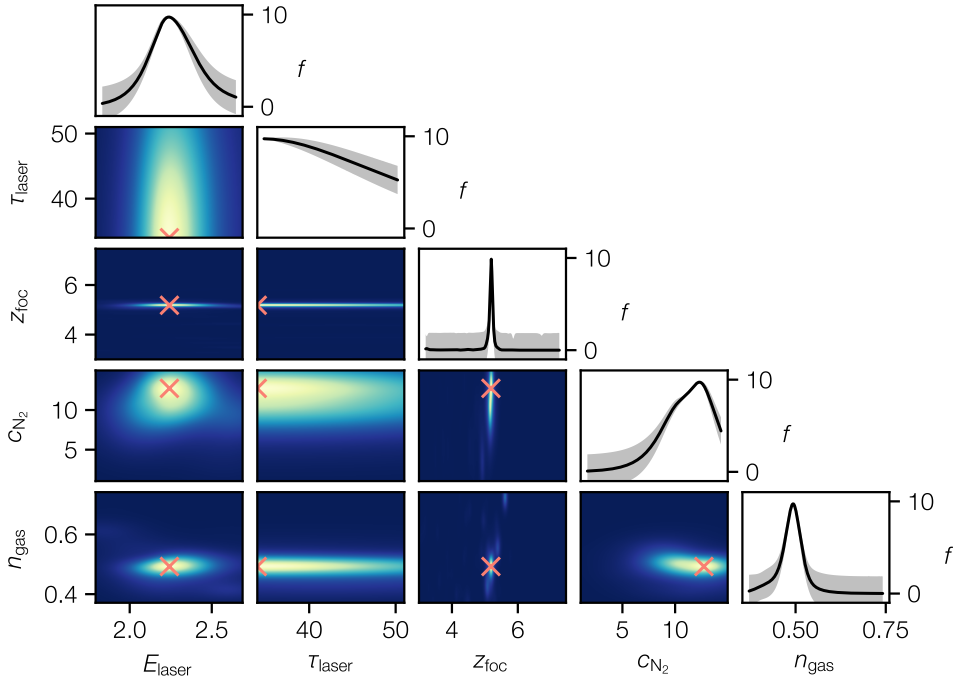

Figure 3.9 – Surrogate model generated from the Bayesian optimization run in section 3.2.3. Shown are the predicted objective function by the model (brighter equals higher) around the optimum for all parameter combination (lower triangle) and scans over the search space with corresponding model confidence for the individual parameters (diagonal). The x indicates the position of the found optimum.

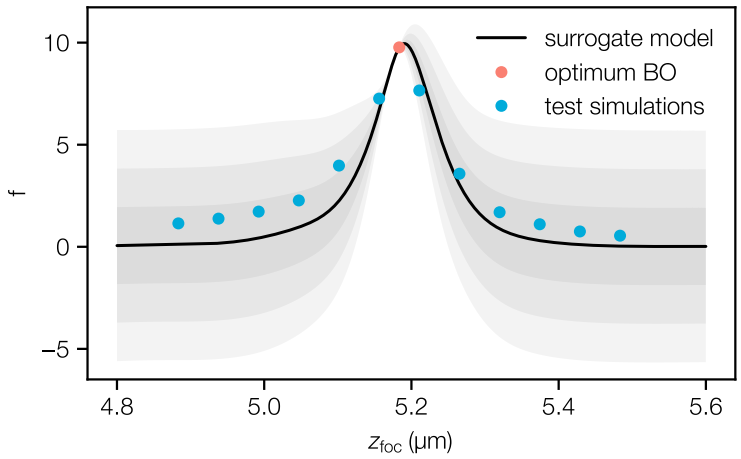

Figure 3.10 – Validation of the surrogate model (black curve, $1, 2, 3 - \sigma$ confidence as shaded areas) around the optimum through additional PIC evaluations (blue dots).

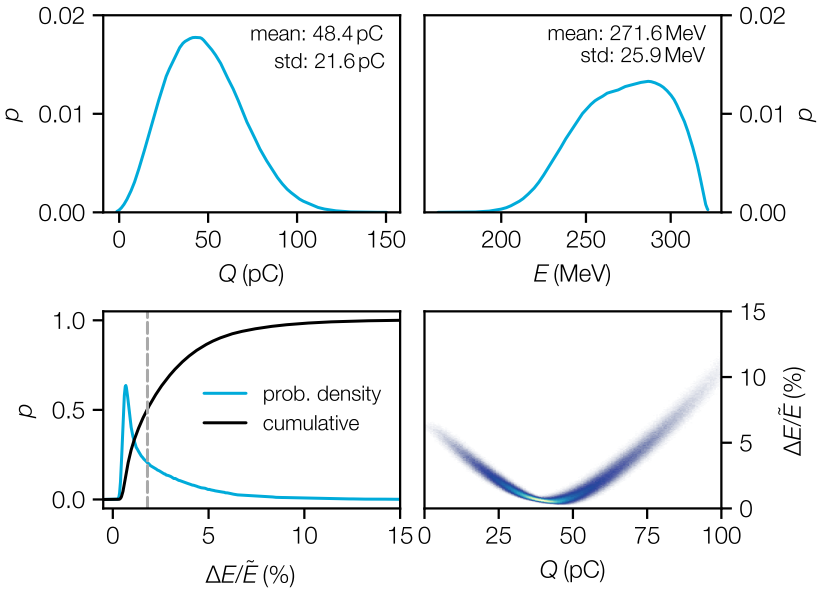

Figure 3.11 – Behavior of the beam parameters when subject to laser fluctuations. The surrogate was evaluated around the optimum with 10^6 samples to resemble jitters in the laser energy, pulse duration and focus position (0.6 %, 2 %, 160 µm rms jitter with independent normal distributions). Shown are the resulting probability density functions for the beam charge (top left), energy (top right), relative energy spread (bottom left) with its cumulative distribution(black line) and median (grey dashed line) as well as the correlation between the charge and the relative energy spread (bottom right).

To test the accuracy of the surrogate model in the direct vicinity of the optimum, an additional scan of the focus position was performed. Figure 3.10 shows the result of these simulations and the predicted objective by the surrogate model. The test simulations are unknown to the model. Although the model slightly underestimates the true objective function far away from the optimum, the true values of the test simulations stay within the uncertainty of the model.

With indication that the surrogate model is reasonably accurate around the optimum it can be used to model the behavior of the electron beam in the presence of realistic fluctuations of the laser system.

Instead of modeling the objective function, the data from the optimization is used to build Gaussian process models for each of the beam parameters directly, i.e., charge, energy and energy spread. The experimental conditions are then simulated by probing the model around the optimum while adding random fluctuations to the laser parameters. These jitters are independent and normally distributed with rms amplitudes assumed based on previous measurements at LUX [5].

Accordingly, the laser energy is assumed with a jitter of 0.6 %, the pulse duration with a 2 % jitter and the focus position with a shot-to-shot variation of 160 µm. Since the cost of evaluating the surrogate model is vanishingly small compared to a PIC simulation it allows to probe

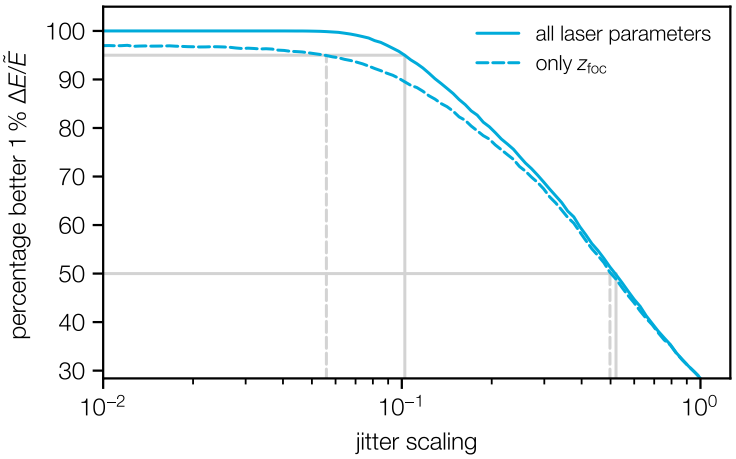

Figure 3.12 – Predicted fraction of shots with less than 1 % energy spread as a function of the laser stability. Grey lines mark the required stability improvement in all laser parameters (solid line) or only the focus position (dashed line) to reach a percentage of 50 or 95.

the system with very large statistics. Here, to model the expected stability of the electron parameters the model is probed with 10^6 samples. Figure 3.11 shows the resulting probability density functions for the electron parameters when subject to realistic shot-to-shot variations of the laser. From the distributions it is immediately clear that the high beam quality that is suggested by the optimization result cannot be maintained for every shot with the assumed laser fluctuations. The analysis suggests that for the beam charge a jitter of 43 % and for the energy of nearly 10 % can be expected. The energy spread has a very asymmetric distribution with a long tail towards high values. For this distribution the rms value is not a very descriptive quantity instead it makes sense to look at properties of the cumulative distribution. Assuming that applications need a certain energy spread to operate, a useful quantity is the fraction of events that result in beams with this spread or better. In this case only 28.4 % of the events give an energy spread of 1 % or better.

The asymmetry of the energy spread distribution is a result of the beam loading effects that are essential to generate the high quality electron bunches. As shown in the bottom right panel in fig. 3.11, the low energy spread, i.e., ideal beam loading, is only achieved for one specific charge. If the charge is higher or lower, i.e., the wake is under- or overloaded, the energy spread deteriorates so that in essence it can only get larger which then gives the asymmetric distribution seen in fig. 3.11.

The stability of the laser assumed for the analysis in fig. 3.11 is the current state of the art in laser plasma acceleration [4, 5, 119] and is mainly limited by low repetition rates of the lasers and therefore missing feedback to stabilize against shot-to-shot fluctuations. With the current advancement toward higher repetition rate systems, a steady improvement in stability is a foreseeable. Therefore, it makes sense to estimate the improvement that is required for the optimized setup to produce beams with a defined shot-to-shot reproducibility.

For this purpose, the jitter of the laser parameters from before was scaled and the surrogate model was again sampled with the improved laser stability. As a figure of merit the percentage of beams with less than 1 % energy spread is used. Figure 3.12 shows scans over the stability of all three laser parameters and only the focus position since this is the most sensitive parameter according to the analysis above.

To go from the initial 28 % of beams that have sub-percent energy spread to 50 % the laser stability needs to be improved by roughly a factor of two. Until this point the jitter of the electron parameters is dominated by the focus position jitter so that an improvement in only z_{foc} yields nearly the same stability as improving all parameters. To improve the electron reproducibility further the other laser parameters become more and more relevant. In order to go to the point where 90 % of the beams have sub-percent energy spread the focus stability alone would have to be improved by nearly a factor of twenty. Improving all laser parameters simultaneously would still necessitate a stability that is ten times superior to the best technology currently available.

3.3 Multi-objective optimization

The previous section's statistical analysis reveals that the beam quality optimum, identified through Bayesian optimization, is greatly sensitive to input parameters, most notably the focus position. While enhancing beam quality remains a core challenge for plasma-based accelerators, the need for improved shot-to-shot stability, a previously often overlooked issue, is equally critical. Foreseeable advancements in laser technology in the coming years will address some of these stability issues. Yet, considering the extreme sensitivity of the optimization result, these advancements alone won't be enough to attain the high stability levels demanded by rigorous applications. Therefore, designing LPAs that are naturally impervious to laser driver fluctuations becomes a necessity.

A straightforward strategy to simultaneously tackle quality and stability challenges could involve incorporating the stability of a setpoint into the objective function. However, the balance between beam quality and stability may vary depending on the application. Some applications might be comfortable with lower beam quality but require higher stability, or vice versa. In the worst case scenario, it could be that higher quality beams inherently possess lower stability with a given setup. In such a case, it's essential to strike the right balance between the two, a balance that best serves the needs of the particular application in question.

As discussed in section 2.3.4, problems like this with multiple potentially competing objectives can also be solved with a Bayesian optimization approach. To do so, instead of choosing points to evaluate based on the expected improvement of a single objective, the acquisition function in this case describes the expected improvement of the hypervolume that is spanned up by all non-dominated solutions. With this the optimizer tries to find points closer and closer to the real Pareto front, i.e., the hypersurface where one objective can only be improved at the cost of another one. Finally, it will provide a set of solutions that are optimal tradeoffs of the competing objectives.

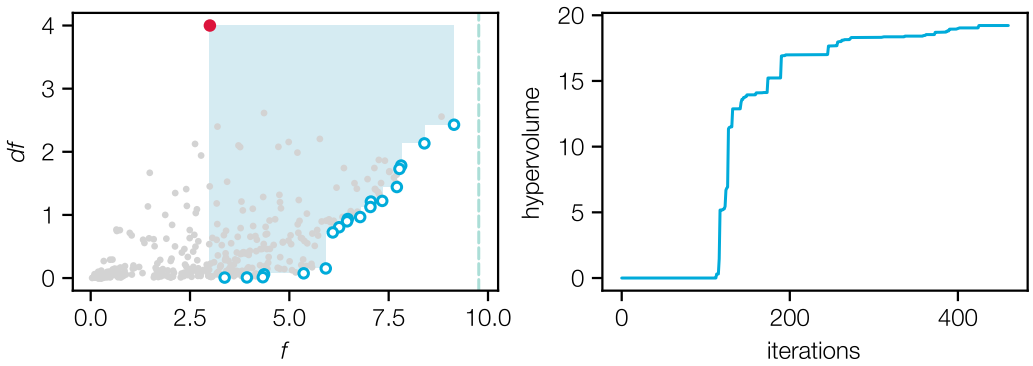

Figure 3.13 – Multi-objective Bayesian optimization of the beam quality f (eq. (3.1)) and its robustness against fluctuations of the focus position df (eq. (3.4)). Left: Simulations results (grey) and the Pareto optimal solutions (blue circles) found during the optimization. Together with the reference point (red) the optimal solutions span a hypervolume (blue area) that acts as the acquisition function for the Bayesian optimization. The dashed line indicates the optimum found with single objective optimization in section 3.2.3. Right: Increase of the hypervolume during the optimization process.

3.3.1 Beam quality and stability tradeoff

To find the optimal tradeoffs between the beam quality characterized by the objective function f in section 3.2 and robustness against laser fluctuations a second objective needs to be defined that describes the stability of a given setpoint. As a simple measure for this the standard deviation of f is used when evaluating a setpoint with varying laser parameters, i.e.,

$$df = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - \bar{f})^2}, \tag{3.4}$$

where $\bar{f}$ is the average of f over all N evaluations.

In practice, evaluating this stability objective including variations of all parameters would require to run many simulations at each iteration of the optimization. Instead, to be more efficient only the most sensitive parameter is varied, i.e., the laser focus position, and only three simulations are run for each optimization step. Accordingly, each setup is simulated at the position suggested by the optimizer and additionally at shifted laser focus positions of $\pm 100\,\mu m$. The beam quality f at the center position and the standard deviation of f are then fed back as objectives.

The multi-objective optimization follows a similar path to the previous single-objective optimization, with three points concurrently computed, but each point now encompasses three simulations with varying focus positions. Upon completion of one set of simulations, the surrogate model gathers the new data, and the acquisition function is maximized accordingly.

As explained in section 2.3.4, for multi-objective Bayesian optimization, the enhancement of the hypervolume, defined by the space between the non-dominated evaluations and a reference point, is employed as the acquisition function. A solution is considered non-dominated if there's no other existing point that outperforms it across all objectives. The reference point outlines the region of interest, directing the optimizer's search for optimal solutions.

This reference point is set somewhat arbitrarily but is typically placed just below the acceptable limit for each objective. In the current case, it's positioned at 3 and 4 for the beam quality objective f and its stability df, respectively. This placement prevents the optimizer from wasting time exploring areas producing either poor quality yet stable beams or the reverse.

Throughout the optimization loop, as illustrated in the right plot of fig. 3.13, the hypervolume incrementally expands. This progression brings about two implications: the discovered solutions gradually converge towards the system's true Pareto front, and the front's coverage improves. This enhancement signifies the identification of more diverse optimal tradeoffs between the objectives.

The final hypervolume, represented as the blue-shaded area in fig. 3.13, demonstrates the results of this multi-objective optimization process.

An examination of the Pareto optimal solutions, depicted as blue circles in fig. 3.13, indicates that no single solution simultaneously optimizes beam quality and stability. It becomes evident that these two objectives are in competition: an enhancement in beam quality f corresponds to an increase in sensitivity df.

Specifically, trying to boost the beam quality factor beyond roughly 6 results in a significant decline in stability. This decline is represented by a jump in the Pareto front and a subsequent rise in the slope between f and df. When the beam quality factor is below this threshold, df is relatively independent of f and remains at a low level.

The point corresponding to the highest beam quality is proximate to the solution identified via single objective optimization, which is represented by the green dashed line in fig. 3.13.

Examining the input parameters of the optimized setups, as displayed in fig. 3.14, offers insight into the abrupt shift in the Pareto front that distinguishes highly stable beams from high quality ones. The figure illustrates the beam quality objective f as it relates to the individual input parameters of the system, with blue circles indicating the points correlating to optimal solutions.

The observed Pareto front doesn't correspond straightforwardly to any single input parameter, indicating that navigating towards optimal solutions involves adjusting all available parameters of the system. However, some patterns emerge in the form of clustering in the input space, providing some explanation for the observed features of the Pareto front.

Notably, points on the Pareto front corresponding to low f values (thus indicating high stability) all feature the lowest nitrogen concentration of 1 % and a focus position close to the mixed gas region. As addressed in section 3.1.3, lower mixed gas concentration results in shorter bunches. Any potential chirp on these beams would have a less dramatic impact on the projected energy spread, leading to less variation of f and therefore a smaller df.

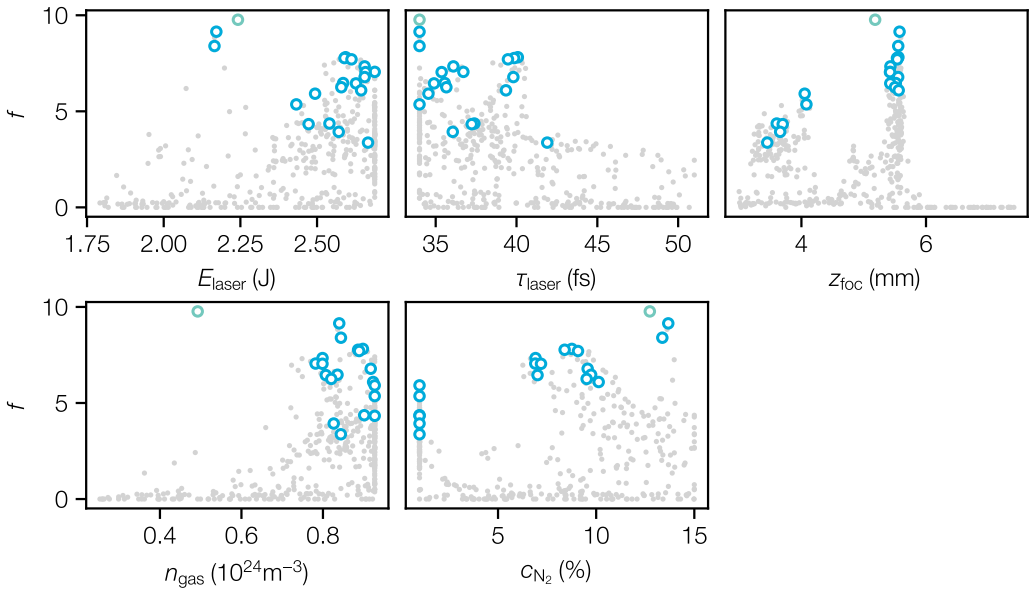

Figure 3.14 – Tested input parameters (grey) plotted against the beam quality objective f and configurations that give optimal tradeoffs between stability and beam quality (blue). The single objective optimum is indicated by the green circle.

However, decreasing the nitrogen concentration inherently reduces the amount of injectable charge. To offset this, the focus must be shifted upstream to augment the ionization and injection volume.

The points with the highest beam quality factor make up another cluster in input space. Here, the focus position is set nearly at the end of the density profile. This allows to significantly increase the nitrogen concentration without injecting too much charge to overload the wakefield. This should result in the injection of long beams with potentially small slice energy spread. If the phase space of such long beams is fully flattened it results in a very narrow energy spectrum with high peak density like the one found with single objective Bayesian optimization in the previous chapter.

As discussed in section 3.1.3, this precise flattening of high order correlations on the longitudinal phase space can be achieved by tuning the laser energy. Consequently, the points on the high quality end of the Pareto front do not use the full laser energy that is available but a slightly reduced one that presumably flattens the phase space. Figure 3.13 also shows the inputs for the single objective optimization before. They are fairly close to the setpoint with the highest beam quality in all parameters except the gas density. As shown in the previous section this point represents a very unstable optimum which therefore makes it extremely challenging to find in the large parameter space. Still, the multi-objective optimizer was able to find an alternative optimum with nearly equivalent beam quality.

To give explicit examples for the presumed effects discussed above, fig. 3.15 shows the electron beams that correspond to the setpoints at each end of the Pareto front and one intermediate

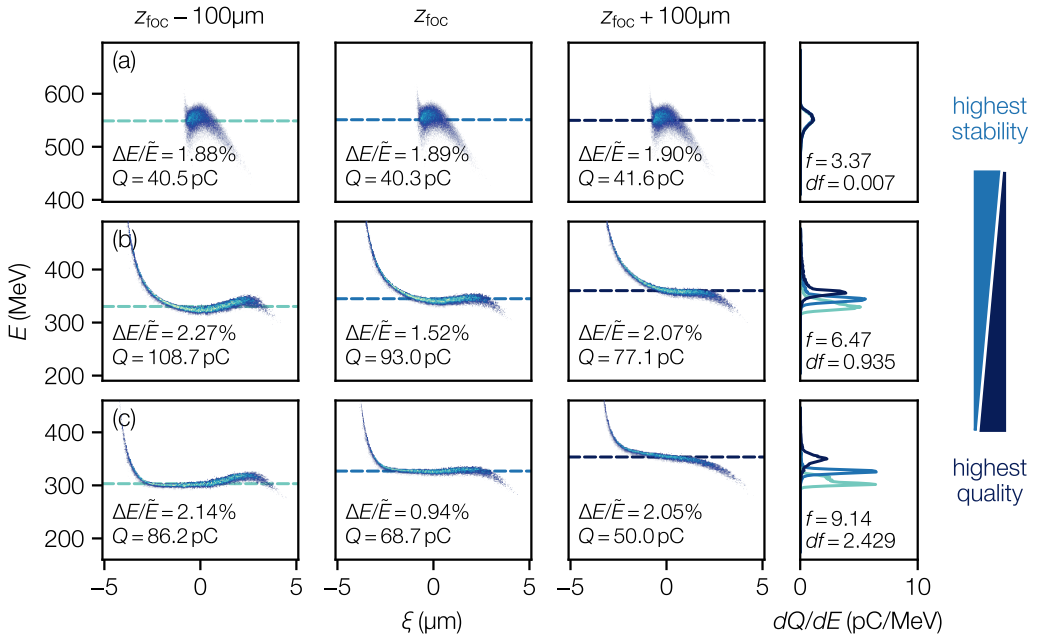

Figure 3.15 – Longitudinal phase spaces and projected energy spectra for different tradeoff between the beam quality objective f and the stability objective df.

tradeoff. The final longitudinal phase spaces are shown for the intended focus position z_{foc} and for a $\pm 100\,\mu m$ offset, i.e., the set of simulations that is used to calculate df during the optimization loop.

The first row shows the configuration with the highest stability. As expected from the low mixed gas concentration, the bunches are very short with a length of $\sim 3\,\mu m$. Additionally, the beam charge is changing only very slightly when the focus position is varied. The combination of short beams with stable beam loading conditions leads to a projected energy spectrum that is nearly unchanged when the focus position shifts. However, since the focus position is closer to the mixed gas region, the injection volume is increased which leads to the large slice energy spread that can be observed. As a consequence the projected energy spread is on the order of 2 %.

At the other end of the spectrum, when stability is entirely sacrificed for the sake of beam quality, as depicted in the bottom row of fig. 3.15, the resulting beams exhibit a bunch length of approximately 8 µm and a significantly reduced slice energy spread. This decreased slice energy spread results from the far downstream focus position and the high nitrogen concentration employed for this setpoint.

With the reduced laser intensity, the injection volume shrinks, and the steep density transition caused by the high nitrogen concentration prompts a rapid change in the wakefield phase where charge is injected. As a result, each longitudinal slice of the beam undergoes only a brief injection event, leading to a small energy spread. Fully flattening this lengthy, narrow phase space reduces the projected energy spread to 0.94 %.

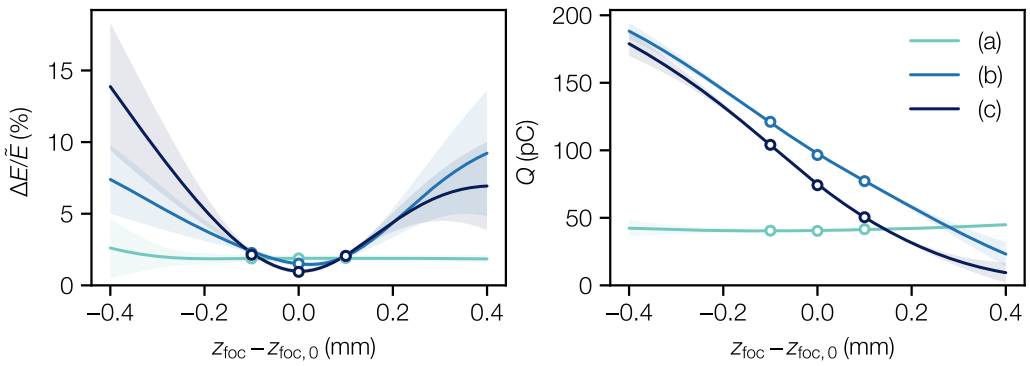

Figure 3.16 – Predictions of the surrogate model for variations of the focus position around the three setpoints in fig. 3.15. Circles indicate the simulated points of each setting and the shaded area marks the three sigma confidence interval of the model predictions.

However, variations in the focus position can cause the beam charge to fluctuate by more than 25 %. This substantial alteration in beam loading imprints a strong chirp on the phase space, and due to the lengthy bunch, it triggers a considerable increase in the projected energy spread. Consequently, the energy spread can surge above approximately 2 %.

The intermediate case in the center row of fig. 3.15 is an example how the objectives can be traded arbitrarily against each other by having knowledge about the Pareto front of the system. The beams here are still changing significantly under variations of the focus position but since they are slightly shorter than the bottom case the relative effect on the projected energy spread is less severe.

Further understanding of the three setpoints can be obtained by utilizing a surrogate model trained on the data gathered during the optimization. Instead of predicting f and df as in the optimizer's case, this model is built to predict the energy spread and the charge. Figure 3.16 presents the predicted properties of the above three setpoints as functions of the laser focus position.

According to the model, all three settings function as local optima for the energy spread, suggesting that the best energy spread is achieved at the design focus position ($z_{\mathrm{foc},0}$). Interestingly, all three settings generate highly similar energy spread for a focus position offset of $\pm 100\,\mu$m. As a result, if the focus position jitter falls within this range, the high-quality setpoint (c) would consistently generate a superior energy spread beam.

However, this phenomenon is primarily a consequence of the parabolic relationship between the energy spread and the focus position, which leads to a small slope around the focus. As one moves further from the design focus position, the three cases begin to diverge significantly. While the energy spread of the stable case (a) remains nearly constant across the entire range, the intermediate case (b) climbs to 7 %, and the high-quality case escalates to nearly 15 % energy spread.

Similarly, the charge for the stable setting remains nearly constant across the full plotted range, whereas the other two settings fluctuate between 0 pC and nearly 200 pC. Although the slope

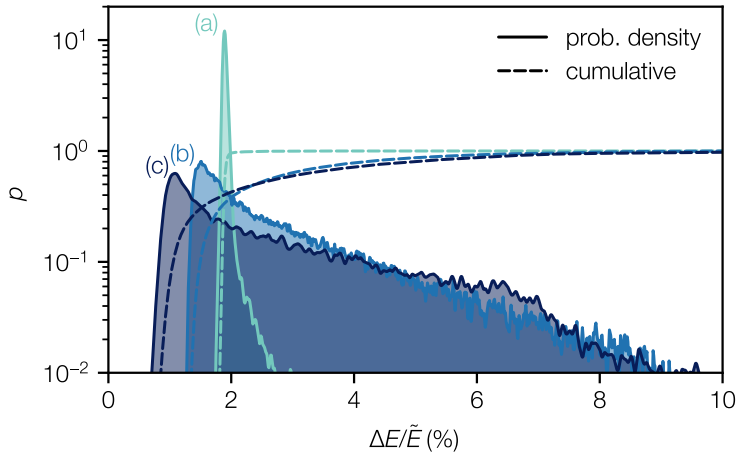

Figure 3.17 – Distributions of the energy spread for the configurations in fig. 3.15 given realistic laser parameter fluctuations.

of the intermediate case is slightly lower than the high-quality case, the charge it produces is marginally higher.

While the goal of this optimization was to identify configurations that remain resilient against variations in the driver laser parameters, the discussion until now has largely centered on the effects of the focus position, based on the presumption that it significantly influences the overall system stability. The advantage of the surrogate model is its ability to simulate the system under any input parameters, thus allowing the examination of realistic fluctuations across all laser parameters.

To compare the three setpoints from fig. 3.15 under realistic conditions 40 000 samples are drawn from the surrogate model for each setpoint with normally distributed random variations of the laser parameters. The rms jitters of the laser energy, pulse duration and focus position are 0.6 %, 2 % and 160 µm, respectively, which is the same as in section 3.2.4 and typical values at LUX.

The resulting distributions of the predicted energy spread are shown in fig. 3.17. Even with fluctuations in all parameters the probability mass of the stable setpoint is extremely localized compared to the other two setpoints. The most frequent energy spreads, i.e., the peaks of the three distributions at 1.9 %, 1.5 % and 1.1 %, still reflect the beam quality difference of the three setpoints. However, due to the long tails of the higher quality settings their mean energy spread is actually significantly larger than that of the stable setting—3 % for setting (b) and 3.2 % for the highest quality setting (c) and still 1.9 % for the more Gaussian distribution of the stable setting (a).

These insights underscore the importance of tailoring the optimal tradeoffs within the system to match specific application requirements. For instance, if an application necessitates the best possible energy spread, using a configuration like (c) is unavoidable. However, for an application with more lenient beam quality constraints—for example, a permissible energy

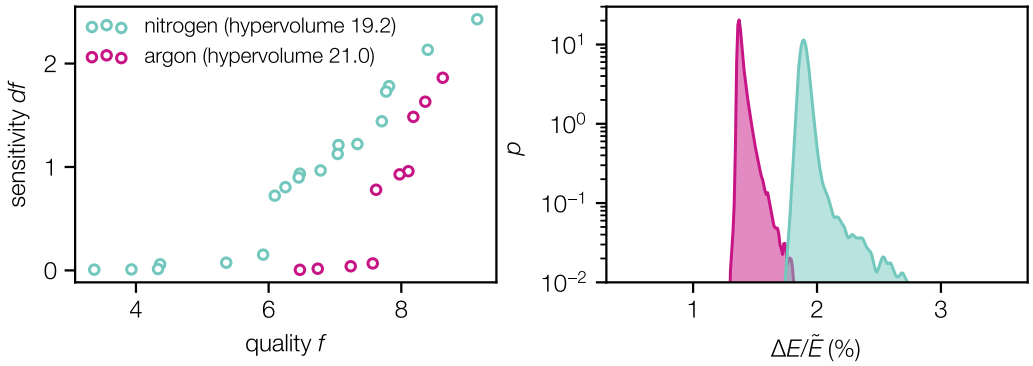

Figure 3.18 – Multi-objective optimization of the setup with nitrogen (green) and with argon (violet) as the dopant gas. Left: Obtained Pareto fronts of the quality objective f and the sensitivity objective df. Right: Energy spread stability of the most stable setpoints for each scenario under the influence of realistic laser jitter. The probability distributions are generated by sampling the surrogate models of the respective optimization runs. Both setpoints have comparable median energy with 550 MeV (nitrogen) and 556 MeV (argon).

spread of less than 2 %—a setup like (a) could actually yield a greater proportion of shots within tolerance, despite a poorer ultimate energy spread. This threshold is delineated by the cumulative distributions in fig. 3.17. Below 2 %, the high-quality settings have a clear advantage as the stable setting simply does not produce such beams. Beyond this point, configuration (c) consistently delivers more shots within the specified tolerance. The cumulative distributions also reveal a relatively marginal advantage of the intermediate setpoint in terms of energy spread stability compared to the high-quality setting.

Finally, an essential question emerges: how could the alignment of these two objectives be improved? More specifically, what modifications could lead to the production of superior quality beams that also display enhanced stability? According to the optimization results, which are considered to approximate the true Pareto front, such an improvement seems impossible within the current setup and parameter space. However, inspection of the input parameters used to generate the Pareto front depicted in fig. 3.14 suggests that the present system might be restricted by the confines of the explored parameter range. Specifically, the inputs leading to optimal solutions—parameters such as gas density, laser energy, and nitrogen concentration—often lie at the limit or closely approach the boundaries of the parameter range.

An alternative to extending the search space would involve fundamentally modifying the current setup—for instance, by changing the dopant gas. However, comparing two distinct setups can pose certain complexities, as each may have specific advantages depending on different aspects. The Pareto front serves a useful role in this context by offering a comprehensive and equitable comparison of the overall capabilities of a setup. This provides valuable insights that could inform choices about potential modifications to the existing system.

Based on the indication that the current setup is potentially limited by the available laser energy, the setup was reoptimized with argon as the dopant gas. The relevant ionization

levels in argon have a lower binding energy and therefore require a lower laser intensity to be accessed.

The side-by-side comparison of the Pareto fronts, achieved with each dopant gas, is displayed in fig. 3.18. This visual comparison reveals that the hypervolume attained with argon exceeds that of nitrogen. Notably, argon is seen to generate beams with similar quality metrics f, but exhibiting superior stability. As a specific example, fig. 3.18 presents the energy spread distributions for the most stable setpoints of each dopant, under the same laser jitter conditions. Remarkably, when argon is utilized, the entire probability mass could be restricted below a 2 % energy spread. However, it's crucial to underscore that the best quality beam realized with argon falls short of the optimum quality accomplished with nitrogen. This observation further emphasizes that the selection of the setup, or in this case the dopant, hinges significantly on the specific beam parameters and tolerances that need to be satisfied. Hence, the choice truly depends on the particular application requirements and operational constraints.

3.3.2 Beam parameter optimization

The previously discussed cases illustrate how Bayesian optimization serves as a tool for identifying optimal tradeoffs in a system with competing objectives, customized for a specific application. Technically, this optimization scenario bridges single and multi-objective optimization: the quality objective f incorporates multiple sub-objectives, including the maximization of charge and beam energy and the minimization of energy spread. Yet, this approach can result in ambiguities in the final outcome. For instance, a combination of smaller energy spread and lower charge or larger energy spread and higher charge can yield identical objective values. As demonstrated earlier, disentangling these objectives into their basic elements and directly optimizing them allows more control. Through understanding the entire Pareto front of all pertinent beam parameters, the system can be fine-tuned to meet the exact operational needs of a given application. Moreover, by understanding the transformation from the input parameter space to the beam parameters defining the Pareto front, dynamic fine-tuning of the beam's properties for optimal results can be achieved.

To explore the versatility of the system in terms of varying the electron beam properties, MOBO is run with four different beam quality metrics as objectives. These are the divergence, emittance, energy spread and charge of the electron beam. The divergence and emittance are averaged over the horizontal and vertical plane.

In addition to these four objectives, a fifth objective is introduced that is the absolute difference of the median energy to a defined reference, i.e., $|\tilde{E} - E_0|$. This objective is added to encourage outcomes in a limited energy range, making the results comparable. In the following E_0 is set to 250 MeV. Combined, the objectives can be summarized by the vector $\mathbf{f}$:

$$\mathbf{f} = \left(\sigma_{x',y'}, \varepsilon_{x,y}, \Delta E, Q, |\tilde{E} - E_0|\right). \tag{3.5}$$

The reference point for the optimized hypervolume is given in table 3.3 and is chosen to promote a baseline of acceptable beam parameters for all objectives.

Table 3.3 – Reference point defining the lower bounds (upper for minimized objectives) of the hypervolume that is maximized in the multi-objective optimization.

	$\sigma_{x',y'}$	$\varepsilon_{x,y}$	ΔE	Q	$\lvert\tilde{E}-E_0\rvert$
REFERENCE POINT	2 mrad	3 mm mrad	10 MeV	10 pC	50 MeV

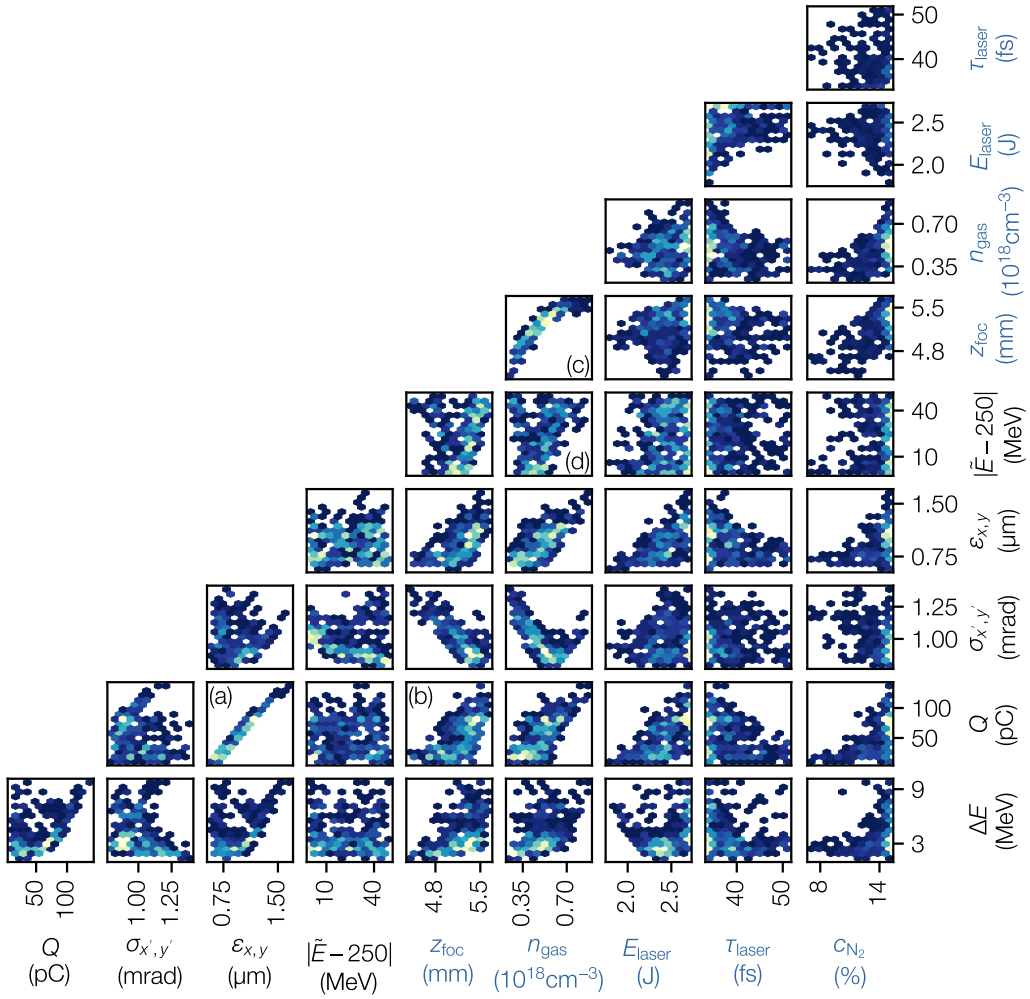

Figure 3.19 – Pairwise correlations of objectives (black) and inputs (blue) of the five dimensional Pareto front.

The optimization process ran for 1000 iterations with three concurrent simulations. The result was a Pareto front comprised of 299 non-dominated solutions. Naturally, comprehending a five-dimensional Pareto front is more complex than the two-dimensional case previously discussed. Yet, a general understanding of the relationship between the objectives and tuning parameters can be deduced by examining their correlations, as shown in fig. 3.19. This figure illustrates the pairwise correlations between the tuning parameters and objectives of the optimal solutions identified during optimization.

Some correlations can be attributed to the basic properties of localized ionization injection discussed earlier, while others are a result of simultaneous tuning of multiple parameters. The most pronounced correlation is between charge and emittance, as seen in fig. 3.19a. This vividly highlights the dilemma of handling conflicting objectives. Any increase in charge invariably leads to an elevation in beam emittance—a compromise that appears inescapable under the conditions of the setup. This is expected when considering that the increased charge is a consequence of a larger injection volume, which also expands the occupied transverse phase space volume of the bunch, hence increasing the emittance.

On the other hand, an unexpected and somewhat counterintuitive relationship is observed between the focus position and charge, as shown in fig. 3.19b. Previous discussions suggested that to adjust the charge using only the longitudinal focus position, the focus should be moved upstream, that is, to smaller values, to intensify the laser in the mixed gas region and subsequently increase the charge. However, in the discovered Pareto front, the focus position is shifted downstream as the charge increases.

This can only be the result of a combined adjustment of laser energy, gas density, and focus position, which collectively create optimal conditions during acceleration. Indeed, the Pareto front explicitly links the focus position to gas density, as evidenced by the clear correlation between the two, illustrated in fig. 3.19c. Considering that increasing the magnitude of these two parameters has contrasting effects (as previously discussed in section 3.1.3), coordinating their adjustments could be a way to finely tune the electron parameters while ensuring ideal beam loading conditions.

The Pareto front has an additional feature that is worth noting, which is the behavior of the beam energy offset relative to the target energy of 250 MeV. When beams are constrained to be close to the reference energy, the gas density is restricted to a certain range. This effect is reflected in the V-shaped distribution shown in fig. 3.19d, which reaches its minimum at approximately $0.45 \times 10^{18} \mathrm{cm}^{-3}$. Furthermore, since the focus position and gas density are linked, a similar V-shaped correlation can be observed between the energy offset and the focus position.

These examples highlight some of the complexities within the Pareto front—intricacies that aren't readily apparent when investigating the system under single objectives and single tuning parameters, such as through dedicated parameter scans. The visualization in fig. 3.19 highlights the most prominent characteristics, but more subtle features, such as the precise interplay between focus position, density, and laser energy, remain obscured due to the simultaneous variation of all parameters. Thus, fig. 3.19 should primarily serve as an illustration of the complexity of the Pareto front in a system with multiple tuning parameters and objectives.

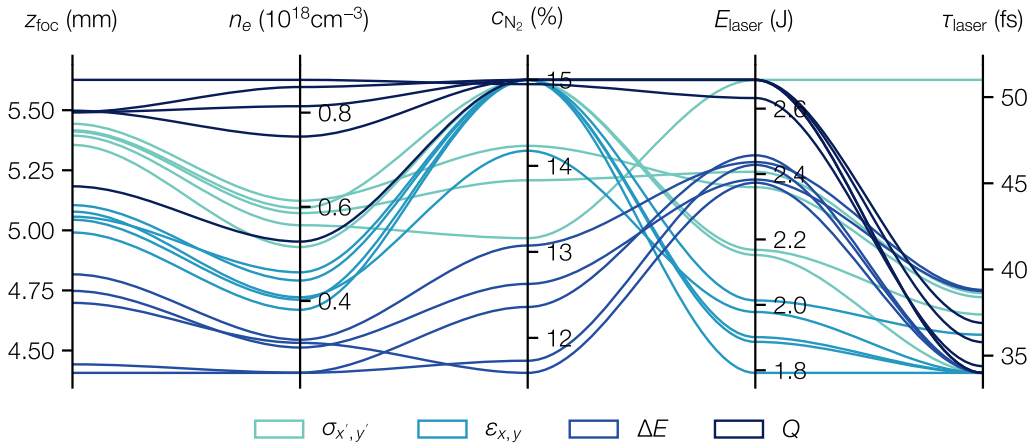

Figure 3.20 – Optimal configurations for each objective presented in a parallel coordinate plot. Each line shows a single optimal setpoint, color-coded by the objective it optimizes. Lines are for guidance and do not imply interpolation. The plot shows that optimal setpoints cluster around specific values for each objective, illustrating the unique requirements of each optimization task.

The impact of the different competing objectives can be observed by examining the regions of the input parameter space that they tend to occupy. As depicted in fig. 3.20, the input parameters of the top five configurations for each beam quality objective exhibit distinct preferences. While pulse duration tends to be towards the lower end of the available range for all but one configuration, each objective demonstrates specific preferences within the available range for the other input parameters.

The distinct usage of the input parameters to optimize each objective naturally leads to different characteristics of the accelerated electron beams. The optima of each objective while keeping the others in the preference region defined by the reference point in table 3.3 are depicted in fig. 3.21. Notably, all objectives produce reasonably flat phase spaces, and only the maximum charge case comes close to the 10 MeV energy spread limit. This confirms that the optimization approach of keeping all objectives as good as possible in all configurations in fact works.

The phase spaces depicted in fig. 3.21 exhibit noteworthy behaviors worth discussing individually. The first phase space showcases the case with the optimal absolute energy spread. The beam has a nearly completely flattened longitudinal phase space with only a very minor tail with a slight energy deviation. Notably, the median energy is on the lower end of the acceptable energy offset at 200 MeV. Further accelerating the beam to the desired 250 MeV might expose it to detrimental effects, such as remaining variations in the accelerating field, potentially causing a slight increase in the absolute energy spread. This could be the rationale for its lower median energy.

The second case focuses on optimizing the beam divergence and still demonstrates an impressively flat phase space. This suggests that energy spread and divergence are relatively

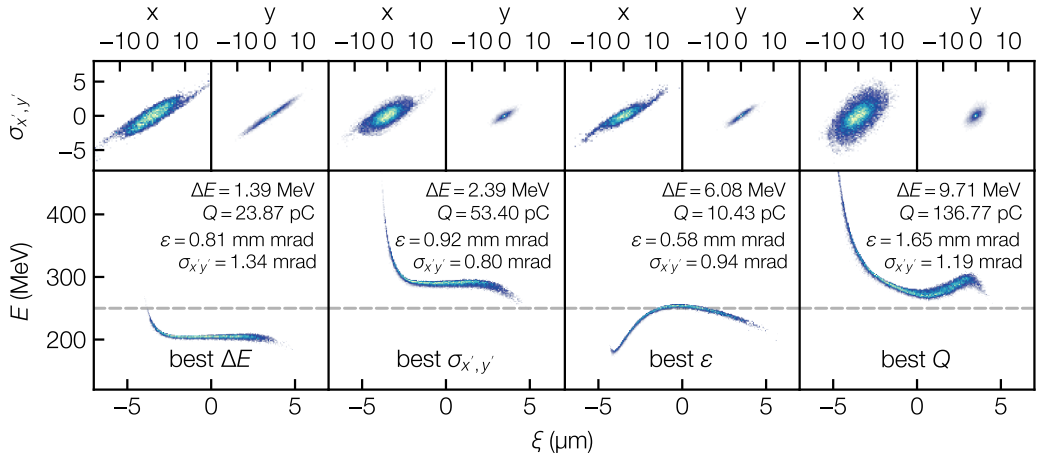

Figure 3.21 – Transverse (top) and longitudinal (bottom) phase spaces for best setpoints for each individual objective. The design energy of 200 MeV is indicated by the dashed line.

compatible objectives. Notably, this beam is at the higher end of the acceptable energy range. This can be explained by the divergence being the ratio of transverse to longitudinal momentum. As the beam accelerates, the divergence naturally decreases, provided that the beam isn't subject to significant deteriorative effects.

The third case, which optimizes for beam emittance, presents a stark trade-off in terms of energy spread and beam charge. When compared to the best energy spread case, a 28 % reduction in emittance results in over a fourfold increase in energy spread. This dramatic increase in energy bandwidth stems from the inability to effectively load the wakefield, manifesting as a correlated energy spread in the longitudinal phase space. Despite these effects, it's worth noting that the slice energy spread in some areas is similar to or even lower than in the first case. This suggests that measures that improve transverse emittance, such as reducing the injection volume, can also enhance the longitudinal momentum distribution.

Finally, the fourth case prioritizes maximizing charge, which inversely leads to nearly a threefold increase in emittance. This high charge beam, which excessively loads the field, features a substantial tail and a nonlinear chirp at the beam's center. This results in the energy spread nearing the acceptable limit of 10 MeV.

These four cases represent the extreme tradeoffs that can be made in the system. In practice, often a much more fine-tuned control over the beam parameters is required. As long as the optimization result samples the Pareto front densely enough, it is possible to custom tailor a beam that matches the requirements of an intended application.

For example, to inject the plasma accelerated beam into a synchrotron typically normalized emittances on the order of tens of mm mrad are sufficient [6]. However, the energy acceptance of these devices is generally very tight so that the LPA needs to provide low bandwidth beams at precisely the design energy. For the sake of this example, it is assumed that the energy

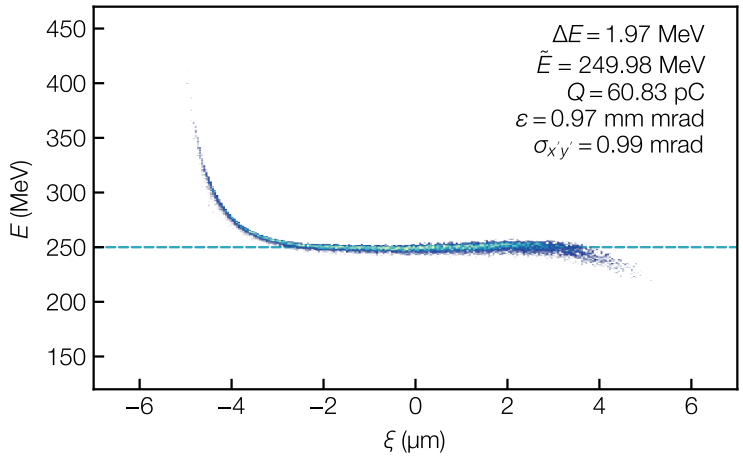

Figure 3.22 – Longitudinal phase space of an electron beam customized to match application constraints. The beam is designed to provide highest beam charge while keeping the deviation from the design energy (dashed line) and the energy spread within a tolerance of 1 %

deviation and energy spread need to be below 1 %[3]. The final free beam parameter is then the beam charge. For efficiency reasons this objective should be maximized while staying within the tolerances for the other parameters. Figure 3.22 shows the beam that best matches these specifications under Pareto-optimal conditions within the optimized parameter space. As required, the median energy of the beam is within 0.01 % of the design energy and the energy spread of 0.8 % is smaller than the specified tolerance. By design of the reference point, table 3.3, the emittance of all considered phase spaces is sufficient. With these constraints the maximum charge that can be loaded into the electron bunch amounts to 60.8 pC.

3.4 Summary

This chapter discussed the application of Bayesian optimization to challenges specific to laser-plasma acceleration. It demonstrated that in a setup like LUX, it is indeed possible to generate beams with subpercent energy spread and substantial charge through a meticulous adjustment of input parameters in a complex and high-dimensional parameter space.

The Gaussian process surrogate model served as a tool in studying these optimized setpoints under conditions with realistic parameter fluctuations like they would be present in the experiment. The analysis revealed that achieving high beam quality incurs a cost in the form of increased sensitivity towards variations in laser parameters, specifically the laser focus position. This sensitivity emerges from variations in the injected beam charge, which

[3] In high brightness synchrotrons the energy acceptance is on the order of 1 % so that the energy spread needs to be significantly smaller than that, which is not achievable with the setup discussed here. Concepts to reach this level of energy control have been proposed in [6, 7].

impacts the energy spread and median energy by altering the beam loading conditions in the wakefield.

Utilizing multi-objective optimization enabled the discovery of trade-offs between the ultimate quality of the beams and their susceptibility to parameter fluctuations. To achieve this, different regimes of the setup were exploited, thus either providing long beams for high spectral density or short beams that are less influenced by changes in beam loading.

Following this, multi-objective optimization was used to manage the key features of the accelerated beams. While the setup does permit finding a good balance among all parameters, trade-offs remain necessary due to inherent competition among some parameters. However, multi-objective optimization results underscore the flexibility of the setup, illustrating its capacity for producing a tailored beam suitable for a specific application.

4 Experimental Setup: The LUX Laser-Plasma Accelerator

The preceding chapter's simulation results demonstrated the suitability of Bayesian optimization for application in laser-plasma accelerators. This potential was put to the test in a series of experiments conducted at the LUX accelerator. This chapter will provide a concise overview of the key instruments and components utilized in these experiments. Section 4.1 will give an overview of the experimental setup, section 4.2 will provide a short description of the laser system. The relevant laser and electron diagnostics will be discussed in section 4.3 and section 4.5. The used plasma sources will be presented in section 4.4. And finally, section 4.6 will give details about the control system.

The experiments presented in the following chapters are building upon the collective efforts of numerous individuals. To delve deeper into specific aspects of the system, readers are encouraged to explore the works of these colleagues. Specifically, references [130] and [131] provide comprehensive discussions on the laser system and diagnostics setup. Extensive coverage on electron beam transport and diagnostics can be found in reference [132]. For a more comprehensive understanding of the plasma source setup and gas dynamics, please refer to references [133] and [69].

4.1 Overview

The experiments outlined in this thesis were performed at the LUX laser-plasma accelerator located at Deutsches Elektronen-Synchrotron (DESY). LUX [12, 4], an acronym for Laser-Driven Undulator X-Ray Source, is an experimental research facility designed to study the generation of stable and reproducible electron beams capable of producing high brightness undulator radiation. A key research emphasis at LUX is the exploration of laser-plasma accelerator based free-electron lasing.

Distinct from many previous LPA experiments, LUX prioritizes a robust machine design that includes extensive online diagnostics and a comprehensive set of control variables. These are tightly integrated into a machine control and data acquisition system.

This design approach enables precise control and feedback throughout the experiment. It covers the generation of the laser pulses, the laser-plasma interaction, and the transport and application of the accelerated electrons. The LUX design paradigm is crucial in enabling the successful application of Bayesian optimization. The rich diagnostic data available allows for the construction of precise surrogate models. These models can then be utilized to optimize the complex parameter space represented by the various tuning parameters. By integrating the optimizer with LUX's comprehensive control system, it becomes possible to achieve fully autonomous tuning and optimization of the accelerator.

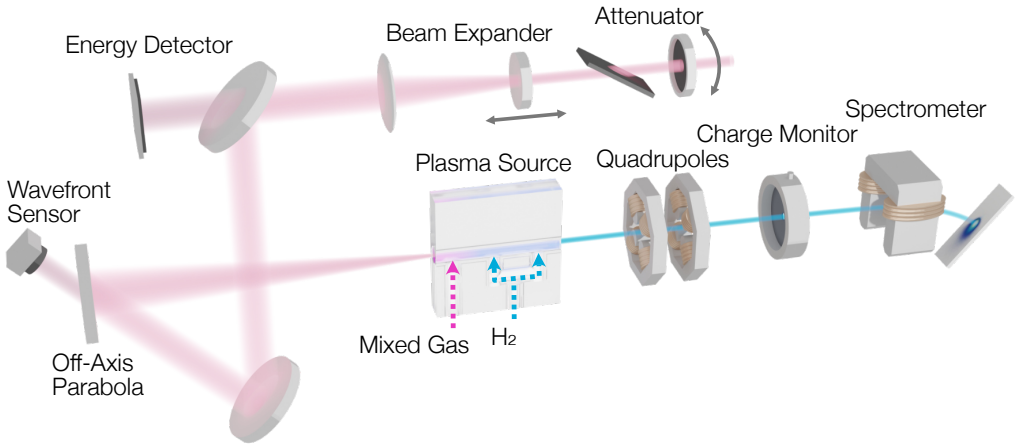

Figure 4.1 – Schematic setup of the experiments discussed in this chapter. The high-power laser is focused into the plasma source, where it drives a plasma wave and accelerates an electron bunch. The accelerated bunch is then focused into a dipole spectrometer by a pair of quadrupole magnets, the charge of the electron bunch is measured using a charge monitor cavity. Figure adapted from [1].

The LUX components can be categorized into three main groups: the high-power laser system, the plasma source, and the electron beam line. The following discussion will focus on the relevant parts of these components for the experiments presented in this thesis.

Figure 4.1 provides a schematic overview of the experimental setup. The plasma interaction is driven by the Ti:sapphire laser system ANGUS, which is focused into the plasma source with an off-axis parabolic (OAP) mirror. The energy of the laser is controlled with a motorized attenuator, operated in closed loop with an energy detector that measures the leaked signal through one of the transport mirrors. The defocus of the wavefront, corresponding to the longitudinal focus position, is measured with a wavefront sensor behind the OAP and tuned by shifting the defocusing lens of the final beam expander in the laser. The plasma source uses localized ionization injection, either with nitrogen or argon, to generate high quality electron bunches. After exiting the plasma source, the diverging electron bunches are captured with a pair of quadrupole magnets and focused into a dipole spectrometer to measure the energy spectrum of the electrons. The bunch charge is measured with a cavity-based charge monitor.

In the upcoming sections, a detailed overview of the LUX setup will be provided.

4.2 ANGUS laser system

To drive the laser-plasma interaction, LUX relies on the ANGUS high-power laser system, which employs chirped pulse amplification (CPA) using titanium-doped sapphire [134] to produce pulses with a wavelength of 800 nm, duration of 25 fs, and energy of up to 6 J.

Originally, the system was designed to operate at a repetition rate of 5 Hz. However, it was discovered that heat-induced deformations of the compressor gratings [11] can reduce the intensity in the laser focus [135]. To prevent this and to also avoid laser-induced damage to the compressor gratings, the system is typically run at a lower repetition rate of 1 Hz, and with a reduced output energy of 3 to 4 J.

ANGUS is comprised of a low-power (∼100 µJ) front-end that runs at a kHz repetition rate and a high-power (∼6 J) back-end running at 1 Hz (originally 5 Hz).

Significant improvements were made to the laser system, particularly to its front-end, around the time of this thesis. However, the results presented in chapter 5 are based on the system's original front-end configuration. During that time, the laser was seeded by a mode-locked oscillator (Venteon) that produced broadband pulses (100 nm) with energy levels in the nano-joule range. These pulses were stretched from a few femtoseconds to more than 100 ps and amplified up to ∼400 µJ in a regenerative amplifier.

After this first amplification stage, the beam was compressed again in order to enhance the temporal pulse contrast through a cross-polarized wave generation (XPW) [136] based filter. After the XPW stage the pulses with ∼40 µJ remaining energy were stretched again for further amplification.

Prior to the measurements described in chapter 6, the original front-end was replaced with a seed laser based on optical parametric chirped pulse amplification (OPCPA) [137, 138]. The system [139] is driven by an industrial ytterbium laser (Pharos-SP) that provides 1 mJ, 1030 nm, 500 fs pulses at 1 kHz repetition rate. A small fraction (1 µJ) of the laser is compressed and focused into a YAG crystal to generate white light in a spectral range of 750 to 850 nm. The remaining fraction of the laser is sent into a β-barium borate (BBO) crystal for second harmonic generation (SHG). The SHG output is used to pump two optical parametric amplifier (OPA) stages that increase the white light energy to a final energy of ∼50 µJ. Again, the pulses are stretched in time to allow further amplification.

Compared to the original front-end, the OPCPA based system provides dramatically improved short-term and long-term stability of the laser properties with measured energy stability as low as 0.2 % over several days of operation [139].

After stretching, the laser pulses are sent into an acousto-optic programmable dispersive filter [140] (Fastlite Dazzler) which allows to control the spectral phase and amplitude. The device allows to tune the final temporal properties of the laser pulses.

Finally, a two-pass Booster amplifier increases the energy to ∼100 µJ before a Pockels cell reduces the repetition rate from 1 kHz to the 1 Hz rate of the back-end.

The back-end features three multi-pass amplifiers that are pumped by frequency-doubled flash-pumped Nd:YAG lasers. These amplifiers increase the energy to first 35 mJ, then 1.4 J, and finally 6 J.

Following the final amplifier, an attenuator comprising a motorized half-waveplate and a pair of thin-film polarizers (TFP) is installed. The waveplate adjusts the laser's polarization, causing a variable fraction of the energy to be rejected by the TFPs. This setup enables stepless tuning of the laser's energy down to approximately 100 mJ. For alignment tasks and low-power measurements an additional pair of neutral density filters can be placed in the beam path to reduce the energy to less than 1 mJ.

Behind the attenuator a Galilean telescope magnifies the beam to ∼70 mm FWHM to reduce the fluence on the subsequent optics. The first lens of this telescope is motorized to slightly tune the divergence of the laser. This proofed to be very useful to precisely adjust the longitudinal focus position.

The magnified beam is sent onto a deformable mirror to compensate wavefront imperfections accumulating in the amplifier chain.

Finally, the beam is sent into an in-vacuum grating compressor to reduce the pulse duration to ∼35 fs. The compressed beam is transported to the accelerator tunnel via a 35 m transport beamline and focused to a spotsize of ∼25 µm FWHM by a 2 m focal length off-axis parabolic (OAP) mirror. Combining the losses from the compressor and the transport beamline, ∼60 % of the pulse energy reaches the plasma source.

4.3 Laser diagnostics

As mentioned at the beginning of this chapter, LUX and ANGUS feature a noteworthy set of diagnostics that allows stringent event based tracing of the experiment from the seed pulses in the laser to the final electron beams. An overview of the most important laser diagnostics used for the measurements in the following chapters is shown in fig. 4.2.

The diagnostic tools for the laser system can be classified as either online or offline. Online diagnostics enable non-destructive measurement of the laser on a shot-to-shot basis, and are designed to operate at full power. They are invaluable for stabilizing the system and analyzing correlations between different parts of the experiment. For example, each stage of the laser system features a full set of online diagnostics to measure the energy, spectrum as well as the near and far field profiles of the beam. The beam profile information is used to stabilize the position and angle of the laser to defined references between each section of the system.

The offline diagnostics require to attenuate the laser to low-power and are used to characterize and optimize the pulse properties prior to experiments. At LUX this is done by inserting a mirror just before the plasma source that reflects the converging beam to a diagnostic table out of vacuum. A 1:1 image of the focus that still lies in the vacuum chamber is projected onto a camera that is mounted on a motorized stage. By moving the stage, the transverse modes of the beam around the focus is measured.

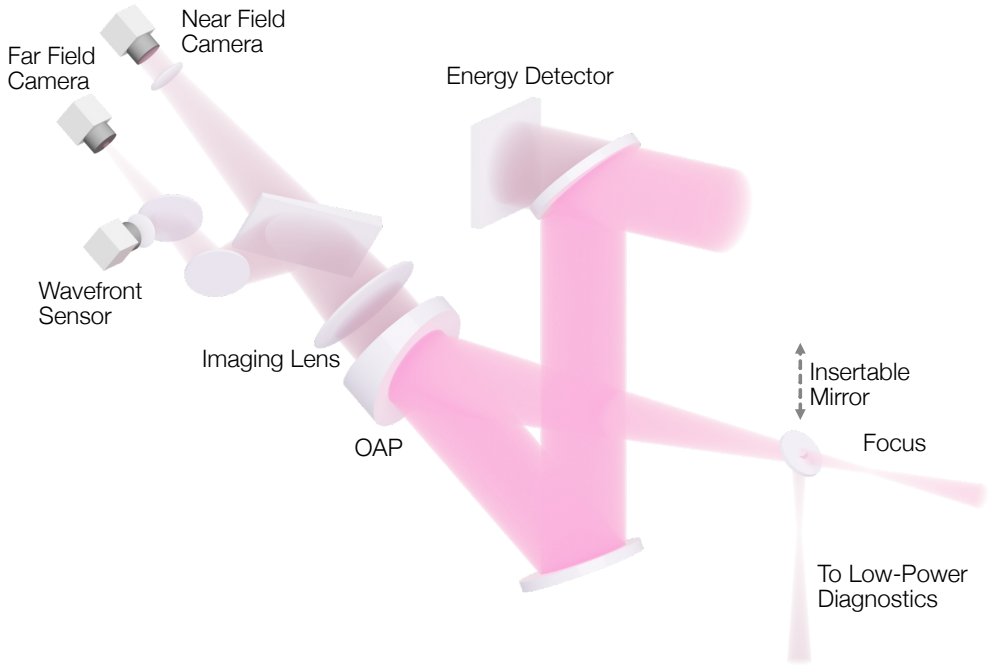

Figure 4.2 – Laser diagnostic setup. The shot-to-shot pulse energy is recorded with a pyroelectric energy detector that measures the leakage through a transport mirror. The leaked signal through the OAP is imaged to a set of diagnostics that measure the near field and far field profiles as well as the wavefront of the incoming pulse. To characterize the beam in low-power mode a mirror is driven into the focused beam that reflects the laser to set of diagnostics outside of the vacuum chamber.

A wavefront sensor [141] is installed in a second diagnostic arm that images the surface of the deformable mirror in the laser system. This wavefront sensor operates in closed loop with the deformable mirror to optimize the focal spot in the plasma source, compensating for systematic wavefront errors such as mirror surface imperfections or misaligned lenses. Running the optimization loop typically achieves a Strehl ratio of > 0.9. An example for the low-power focus measurement is shown in fig. 4.3.

The low-power diagnostic has a third arm to measure and optimize the temporal properties of the laser. The corresponding measurement is done through self-referenced spectral interferometry (SRSI) [142]. To achieve ultra-short pulses, a closed-loop procedure is implemented between the measurement device and the acousto-optic programmable dispersive filter in the laser front-end to flatten the spectral phase. With this method, pulse durations of 30 to 40 fs can be reached. An exemplary measurement is shown in fig. 4.4.

During high-power operation it is essential to measure the properties of the laser non-invasively to ensure an undisturbed delivery of the pulses to the experiment. This is achieved by capturing the leaked laser signal through various optics in the system. A pyroelectric detector measures the shot-to-shot energy behind the last mirror before the pulse compressor. To obtain absolute measurements, the detector is calibrated against a larger pyroelectric detector placed directly

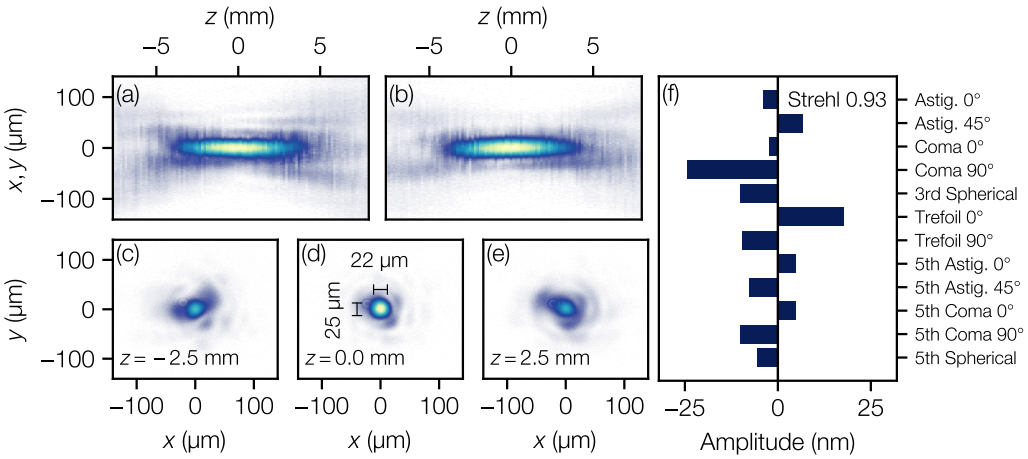

Figure 4.3 – This figure presents the measurement of the low-power focus. Panel (a) and (b) show the horizontal and vertical cross-sections of the beam profile at varying z-positions around the focus, respectively. Panels (c) to (e) display the transverse profiles of the laser around the focus point, along with the Full Width at Half Maximum (FWHM) of the focal spot. Panel (f) illustrates the residual wavefront aberrations post optimization with the deformable mirror.

into the fully amplified beam before the experiments. One challenge with measuring the beam energy behind a mirror is the significant spectral dependence of the transmitted signal through the highly reflective coatings of the optics. As a result, a drift in the pulse's central wavelength could be incorrectly interpreted as a change in the energy. As a spectrally independent energy diagnostic, an additional pyroelectric detector is placed in the reflected beam of the polarizers in the attenuator. However, this measurement is inversely dependent on the setting of the waveplate and is therefore difficult to use in a closed loop with the attenuator, when the goal is not only to stabilize, but also to tune the pulse energy.

The previous chapter demonstrated that the longitudinal focus position plays a critical role for the performance of the LPA. Variations in the laser divergence can cause unintended shifts in the focal plane. During high-power operation, thermal distortions in various optics, particularly the compressor gratings [11], and the mirrors that transport the beam can cause the wavefront to drift away from its optimized low-power state.

To detect deteriorative effects on all optics, a wavefront diagnostic must be positioned as close to the focal point as feasible. As shown in fig. 4.2, a final laser diagnostic setup is installed behind the OAP. A lens behind the OAP collects the transmitted signal, which is divided into three beam paths after exiting the vacuum chamber. The first arm collimates the beam to image its near field in the plane of the OAP surface. In the second arm, a camera measures the focused signal to detect the beam's far field, i.e., the angular distribution on the OAP. The centroids in the near and far field correspond to the angle through the plasma source and the transverse position of the focus, respectively.

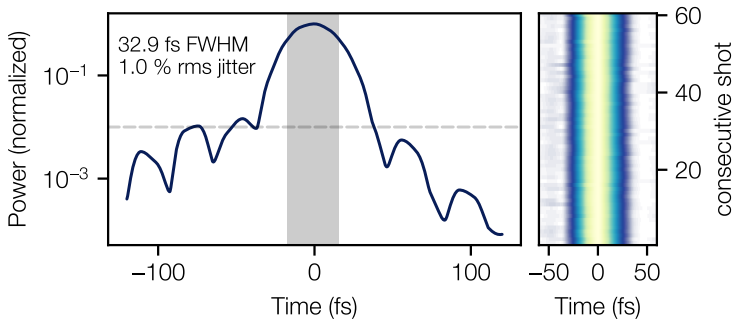

Figure 4.4 – The measurement of the pulse duration with SRSI post spectral phase optimization is illustrated. The left panel shows the average temporal profile over 60 shots, featuring an average pulse duration of 33 fs (FWHM, represented by the grey area). Notably, post-pulses fall below 1 % of the peak power (delineated by the dashed line). The right panel presents the temporal profiles of consecutive individual shots. The pulse duration has an rms jitter of 1 %.

A wavefront sensor is located in the third arm of the diagnostic, which also images the OAP surface. Although this sensor cannot measure the absolute wavefront in the vacuum chamber due to the thick OAP substrate, it can detect changes to the initial state caused by thermal effects. Crucially, this diagnostic can detect drifts in the beam divergence that cause a shift in the longitudinal focus position [130].

To get a calibration for the actual focus shift from a change in the divergence, the laser was focused into the vacuum chamber that hosts the plasma source, which was filled with hydrogen. The recombination signal of the plasma that was created around the focus was used to determine the focus position [133, 69]. By shifting the focus with the movable lens in the beam expansion telescope a calibration for the defocus component of the wavefront was recorded. According to this, a change of the defocus (rms normalization) of 1 nm corresponds to a 26 µm shift of the longitudinal focus position. The typical jitter of the defocus at LUX is on the order of 5 to 10 nm rms, i.e., a focus position jitter of 130 to 260 µm.

4.4 Plasma source

The general concept of the localized ionization injection based plasma source at LUX was already discussed in section 3.1. In practice, the plasma source concept is realized by milling the geometry into a pair of $10 \times 10 \times 3$ mm sapphire slabs, which are then mounted against each other in a sandwich-like configuration. The plasma source design was originally conceived by Manuel Kirchen and discussed in [69] and [5]. The experimental results presented in chapter 5 are based on this original version, while an improved design was implemented at LUX prior to the measurements presented in chapter 6. The structured halves of the two plasma sources is shown in fig. 4.5.

Both sources operate based on the same underlying concept. As explained in section 3.1, mixed gas and pure hydrogen are supplied through separate inlets. These gases do not fully

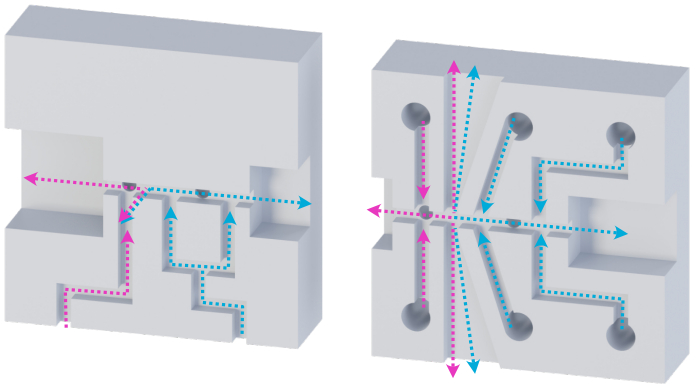

Figure 4.5 – Geometries of the original plasma source used in chapter 5 (left) and the improved version used for the experiments in chapter 6 (right). A second sapphire slab (not shown) is pressed against the machined sapphire to seal the channel structure. The plasma sources are supplied with a mixture of hydrogen and nitrogen/argon (magenta) and pure hydrogen (blue).

dissolve but instead form a short transition region, approximately 500 μm in length [69], where electron injection takes place.

To minimize the length of this transition region, a perpendicular exhaust channel is incorporated alongside the axial channel responsible for the laser-plasma interaction. This arrangement allows the two gas species to escape quickly instead of extensively mixing with each other. In the original design, the exhaust port was drilled into the cover plate. However, in the new design, two triangular channels are machined into the main sapphire plate. This symmetric design ensures that the transition from mixed to pure gas occurs perpendicular to the laser axis, eliminating potential unintended laser perturbations caused by a tilted transition region.

Furthermore, compared to the single-sided inlets in the old design, the new design features inlets from both the top and bottom, further enhancing the symmetry. In both designs, two additional lateral accesses allow to measure the on axis pressure in the mixed gas region and in the plateau with two capacitive pressure gauges.

The gas flow through the inlets is managed using dedicated mass flow controllers (Bronkhorst EL-FLOW) for each gas species and inlet. In the original configuration, three controllers are employed to regulate the flow through two inlets. Two controllers control the flow of hydrogen and the dopant gas in the mixed gas inlet, while an additional controller manages the hydrogen flow in a second inlet that is divided to create the gas plateau used for acceleration. In the updated design, individual controllers are employed at each end of the plateau section, enabling enhanced flexibility, such as the creation of density ramps in this region.

Figure 4.6 shows the results of computational fluid simulations (Ansys Fluent, see [69] for details) of the two plasma source designs, in which the gas flows are tuned to roughly create the same on axis density in both cases. As shown, both sources generate similar gas density

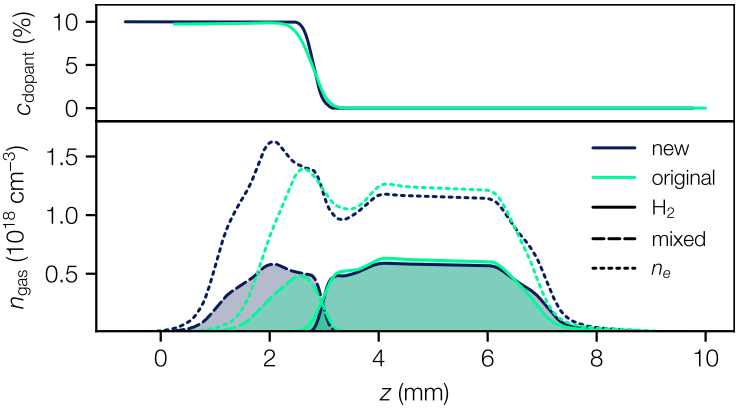

Figure 4.6 – Fluid Simulation of the original (green) and updated (blue) plasma source geometries. Both sources are supplied with a 1:9 mixture of nitrogen and hydrogen (represented by dashed lines) in the first inlet, while pure hydrogen (denoted by solid lines) is supplied in the plateau region. In the transition region, as depicted in the top plot, the dopant concentration rapidly decreases to zero, thereby enabling localized ionization injection. The dotted lines indicate the plasma density after the ionization of the molecular hydrogen (2 electrons) and the outer shells of the nitrogen (10 electrons).

profiles. In the new design the exhaust outlet is positioned slightly further inside the plasma source, giving a broader mixed gas region. However, by having exhaust outlets on each side of the central channel, more gas can escape which leads to a steeper transition between the mixed gas and the hydrogen (see upper plot in fig. 4.6).

In the experiments described in chapter 5, nitrogen was employed as the dopant gas, while argon was used for the experiments discussed in chapter 6. The relevant levels of argon have a lower ionization potential and therefore allow for operation at lower laser energy. This approach has since been adopted to prevent damage to the compressor gratings and, as shown in the previous chapter, the use of argon can result in more stable setpoints.

4.5 Electron beam transport and diagnostics

A distinguishing feature of LUX are the state-of-the-art beam optics and diagnostic components that are also used for the conventional machines at DESY. These allow advanced manipulation and accurate measurements of the phase space of the accelerated electrons. For the FEL experiments at LUX, 13 Quadrupole magnets, 10 corrector dipoles, a decompression chicane and various beam position monitors (BPM) [143] and profile screens can be utilized to match the electron beams into the undulator. For the experiments presented in this thesis only a small subset of these is used to diagnose the relevant properties of the bunches at different machine setpoints.

An overview of the utilized beamline elements is shown in fig. 4.7. The diverging electron beams are captured by a quadrupole doublet (Q3, Q4) positioned 3.9 m downstream of the

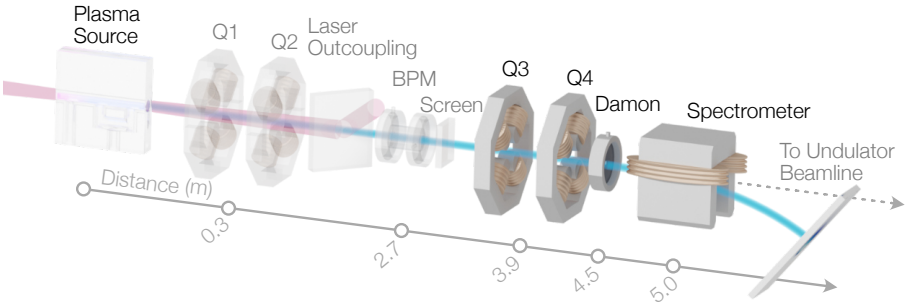

Figure 4.7 – Sketch of the electron beamline. The electron beam (blue) at the exit of the plasma source is imaged onto the fluorescent screen in the spectrometer by the quadrupole doublet Q3/Q4. A cavity (Damon) measures the bunch charge. Only the first section of the LUX electron beamline is shown; components that are not explicitly used in this chapter are greyed out.

plasma source. These quadrupole magnets image the electron beams from the plasma exit onto a fluorescent screen (Kodak Lanex OG 16) located behind a deflecting dipole magnet [144]. Due to the energy dependent deflection angle this acts as a spectrometer for the electrons.

Prior to entering the spectrometer, the electrons pass through a resonator cavity (Damon) [145] for a bunch charge measurement. The Damon measurement works by detecting the TM_{01} mode induced by the electron bunch in the resonator, with the voltage of the signal directly proportional to the beam charge. The Damon has a measurement resolution of approximately 1 % [145].

The simulations conducted in the previous chapter suggest that the LPA setup has the potential to achieve energy spreads as low as 0.5 %. To fully investigate these capabilities in the experiment, the diagnostics setup must obviously be capable of providing the corresponding energy resolution. However, if the electron bunches were allowed to propagate freely to the spectrometer, the typical divergences of 1 mrad [69] would result in beam sizes of approximately 5 mm on the screen, significantly constraining the achievable resolution. By employing a suited beam optic that images the plasma exit into the spectrometer, the resolution at a desired energy can be severely increased. To demonstrate this, beam tracking simulations using the transport matrix based code LUX TRACKER [146] are performed. Figure 4.8 shows the simulated signal on the spectrometer when tracking monoenergetic beams at different energies through the beamline that was optimized to image 250 MeV in this example.

Due to the large distance between the quadrupoles and the plasma source (3.9 m) in comparison to the spectrometer (1.1 m), the beam is demagnified in the spectrometer. This demagnification results in a focused spot size of <1 μm for a typical source size of 4 μm, theoretically providing an energy resolution well below 0.1 %. However, the optical imaging system, granularity of the scintillator screen and scattering in the wall of the vacuum chamber before reaching the screen, set a lower limit for the spatial resolution at ∼200 μm, which on the energy axis is equivalent to ∼0.1 %. Additionally, as the beam deviates from the imaged energy, it either focuses in front of or behind the screen, resulting in an increase in beam size and the

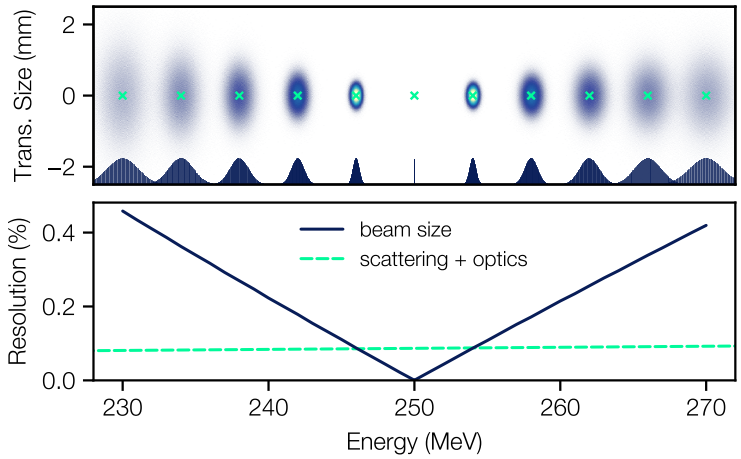

Figure 4.8 – Estimation of the electron spectrometer resolution for 250 MeV focused energy. (Top) Beam tracking simulations of monoenergetic beams at different energies (green crosses) with 4 μm source size and 1 mrad divergence. Blue curves indicate the projection onto the energy axis, the in-focus beam (250 MeV) is too small to be resolved in the plot. (Bottom) Energy resolution around the focused energy given by the rms beam size along the energy axis (blue, solid) and by the optical system and scattering effects in the vacuum chamber (green, dashed).

characteristic butterfly shape on the spectrometer screen and a consequent degradation in energy resolution.

Overall, the spectrometer combined with the beam optics setup provides the required resolution of <0.5 % over a range of ±20 MeV around the imaged energy (see fig. 4.8 bottom). Similar, resolutions are achieved for focused energies in a range of 200 to 300 MeV

To avoid chromatic growth of the beam emittance, that results from free drift propagation of the diverging beam [147, 148, 132], LUX is equipped with an additional pair of quadrupoles positioned directly behind the plasma source. These magnets also in principle allow for less tight focusing into the spectrometer, resulting in smaller beams over a wider energy range. However, the high magnification of this setup makes the measurement more sensitive to variations in the source position. Therefore, this configuration was not utilized for the optimization experiments in the following chapters, as the source position could vary significantly from one setpoint to another.

4.6 Control system

To leverage the full potential of all components presented in the previous sections, a machine control system is tightly integrated into LUX. The system, which is based on the DOOCS framework [149], allows real time access to all diagnostics in the system and control of every actuator, e.g., motors, flow controllers or magnets. A data acquisition system (DAQ) records

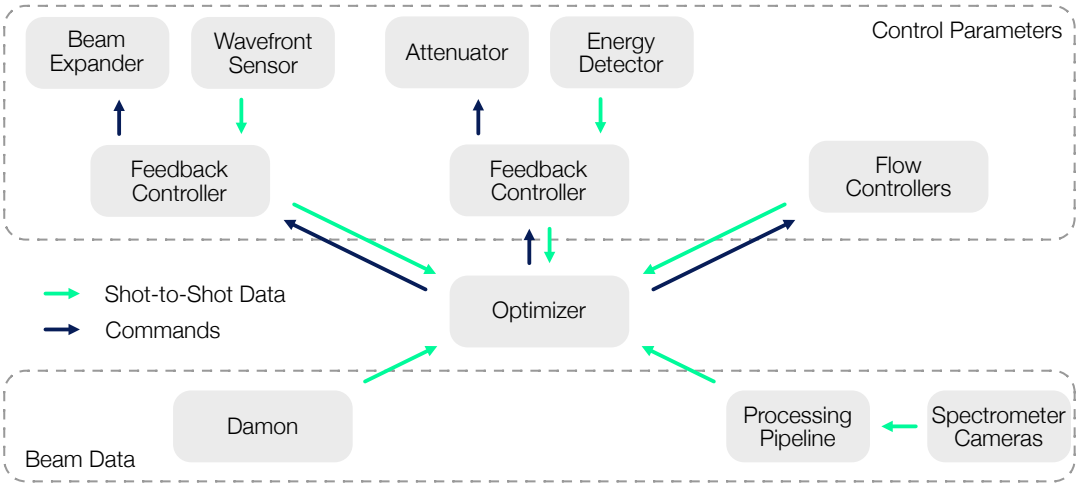

Figure 4.9 – Control system schematic for Bayesian optimization at LUX. The optimizer sets the target values for feedback controllers that tune the machine to the specified defocus and energy. The flow controllers provide an internal control loop. The optimizer gathers data from the electron diagnostics that is processed in real time through data processing pipelines.

the complete state of the machine for every individual shot and allows data driven post-analysis of the experiments [4, 5].

The robust control system implemented at LUX facilitates the effortless integration of diverse control algorithms necessary for autonomous operation and optimization of the machine. In this context, the Bayesian optimizers employed for this thesis could be seamlessly incorporated as plug-ins within LUX's existing data and command framework. A schematic overview of the setup is shown in fig. 4.9. As in the previous chapter, the optimizer is allowed to control the focus position, via the wavefront curvature, and energy of the laser as well as the various gas flows in the plasma source. As feedback it gathers the beam charge measurement from the Damon and spectral data from the spectrometer. The raw data from the cameras recording the spectrometer screen is run through a processing pipeline [150] to retrieve the denoised, calibrated spectrum in real time. For more detailed information regarding the processing steps, refer to [132].

As indicated in fig. 4.9, the optimizer does not directly communicate with the beam expander and the attenuator, which in essence both are just stepper motor stages moving the respective optics. Instead, the communication is run through a feedback controller. In principle, the Bayesian optimizer could use the motor positions of the actuators as inputs, but this poses certain downsides. Most importantly, the state of the system might drift independently of the motor position, eventually leading to a misinformed surrogate model. Further, the laser parameters undergo shot-to-shot variations. If motor positions are used as inputs, the optimizer would not be able to detect these variations.

The feedback controllers allow to specify the desired physical state of the system in terms of the focus and laser energy directly and potential drifts of the laser will be caught and compensated.

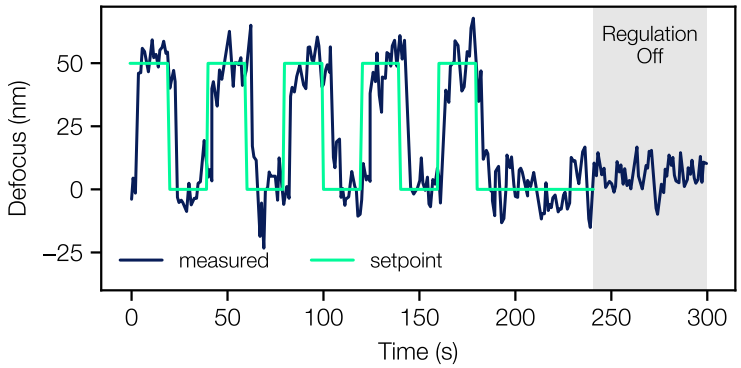

Figure 4.10 – Regulation of the laser defocus using the motorized lens in the final beam expander. The feedback control system utilizes the wavefront measurements taken behind the OAP. The measured defocus (blue) promptly tracks the sequence of setpoints (green).

Having the physically relevant properties as control variables also enables the utilization of individual shot-to-shot measurements for training the surrogate model to account for the laser jitter.

To regulate the focus position, the first lens in the last telescope of the laser system (beam expander), which is placed on a motorized stage, is shifted longitudinally. Feedback for this movement is based on the defocus measurements of the wavefront sensor that is placed behind the OAP. The controller is a simple P-type feedback loop, i.e., it adjusts the control variables based on the difference between the desired state and the current state, using proportional gain to determine the correction. Figure 4.10 shows an example where the feedback controller is used to cycle the system between two defocus setpoints. The gain of the system is adjusted to strike a balance between achieving a quick response and minimizing overshooting. This tradeoff ensures that the system rapidly reaches the desired setpoint without exceeding it significantly, avoiding any potential damage caused by setting the focus too far away from the plasma source. The controller effectively tunes the defocus to the target state within a few seconds.

However, around the setpoints, there are significant shot-to-shot variations, that the control loop does not compensate. They have a comparable amplitude to the unregulated system as shown in the last interval in fig. 4.10 where the regulation is turned off, meaning that they are an inherent property of the laser and not caused by the regulation, e.g, by introducing high frequency noise.

A detailed analysis of the regulation bandwidth is shown in fig. 4.11. For the analysis two ∼50 min timeseries of the defocus were recorded; one in free drift and the other with feedback regulation. The time domain graph already shows that the free running system undergoes slow drifts on the order of a few nm on a timescale of several minutes. These drifts are effectively suppressed in the regulated case. The short term, shot-to-shot variation of the defocus error is comparable for the two data sets. The frequency domain analysis shows that the feedback controller successfully mitigates almost all drift components below 0.01 Hz. Overall, the regulation reduces the mean absolute defocus error from 6.08 nm to 5.88 nm. However, the

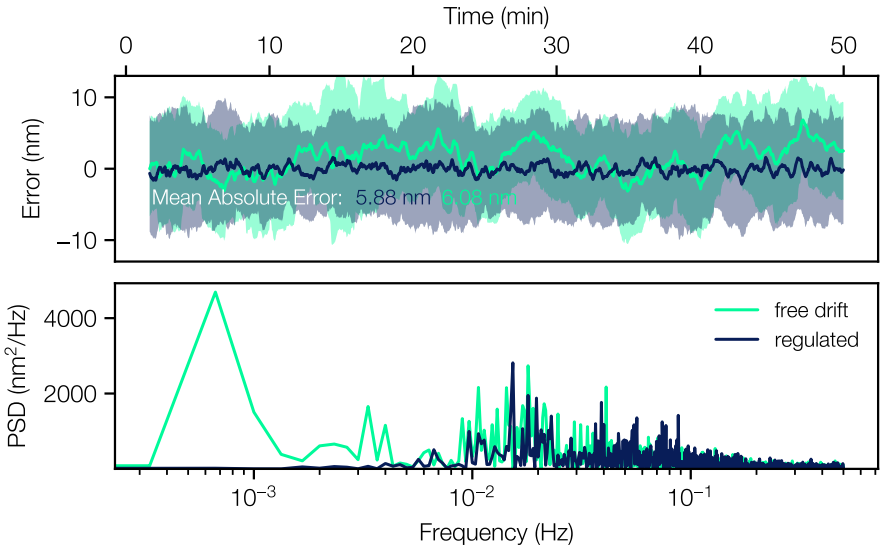

Figure 4.11 – Performance of the defocus regulation. (Top) Time domain signal of the defocus error, i.e., difference between target and actual defocus, with regulation on (blue) and off (green). The lines indicate the rolling average over 100 s and the shaded areas ± one standard deviation around the average. (Bottom) Power spectral density of both measurements. Low frequency jitters are effectively suppressed by the regulation.

length of these sample dataset is still comparably short compared to typical measurement campaigns. Especially, during warmup of the system or during larger temperature shifts from day to night, the defocus drift can be more severe [4] and the regulation therefore more effective. Also note that the feedback controller was specifically tuned for the task of quickly cycling through different setpoints and not primarily for jitter mitigation. While the controller bandwidth is ultimately limited by the 1 Hz laser repetition rate, further improvements could be achieved by using a well tuned PID-controller instead of the simple P-controller employed here.

The same feedback controller method is used for the laser energy. Here, the motorized waveplate of the attenuator is moved with feedback of the energy detector that measures the leakage through the last mirror before the compressor. The mass flow controllers are equipped with internal PID controllers.

5 Bayesian Optimization of a Laser-Plasma Accelerator

This chapter presents experimental results on the application of Bayesian optimization for the improvement of beam quality and stability at LUX. The discussion primarily builds on findings published in the work *"S. Jalas et al., Bayesian Optimization of a Laser-Plasma Accelerator, Physical Review Letters (2021)* [1]. Although this chapter mostly mirrors the contents of the publication, it also incorporates additional details to offer a more comprehensive discussion of the topic.

5.1 Overview

Laser-plasma accelerators have demonstrated tremendous potential in previous experiments, showcasing the generation of few-femtosecond [46, 45] GeV-level electron beams over centimeter-scale distances [53, 56].

However, despite the remarkable progress in the field, meeting the demands for high-quality electron beams necessary for practical applications [4, 151, 152] remains a significant challenge. This difficulty arises from the need to balance various physical effects that non-linearly interact with laser and plasma parameters. The process involves generating electrons with a small initial phase space volume and then accelerating them without compromising the beam quality. Also, for efficient operation, the wakefield must be properly beam-loaded. Achieving these optimal conditions is only feasible in a tiny fraction of the vast parameter space of the system, rendering manual tuning almost impossible.

Additionally, investigating the underlying mechanisms of laser-plasma accelerators is often associated with substantial effort. The limited repetition rates and poor shot-to-shot stability of current drive laser systems, combined with measurement noise, impose constraints on the quantity and quality of data that can be collected about the system.

In this context, Bayesian optimization, designed to find optimal solutions based on limited and noisy data [96], is a promising approach to optimize the complex parameter space of particle accelerators.

As demonstrated in chapter 3, Bayesian optimization is well suited for the application in LPA setups. However, the discussion there was limited on the application on particle-in-cell simulations. While these simulations benefit from the cost-effectiveness of Bayesian optimization, they present an idealized view concerning measurement noise and input parameter variations.

Therefore, translating these results to the practical experiment poses two major complications:

- The input and output parameters are partly obfuscated with significant measurement noise.
- System parameters might deviate from the machine's design configuration requested by the optimizer due to shot-to-shot fluctuations.

The influence of these factors on the performance of the BO algorithm is discussed in section 5.2.

By successfully applying BO in the experiment at LUX, we optimized the accelerator to produce sub-percent energy spread beams at 254 MeV and few-10 pC bunch charge. The details of these optimization results are presented later in this chapter in section 5.3. Furthermore, a slight adjustment to the BO algorithm allowed us to identify a regime that improves the stability of these beams. This particular aspect is further discussed in section 5.4. The physical characteristics of the optimized setpoints will be discussed in section 5.5.

5.2 Parameter fluctuations and noise

Experimental conditions in laser-plasma accelerators can give rise to several forms of noise. This noise can originate from measurement uncertainties in the diagnostic tools, or from fluctuations in unmonitored laser parameters, which ultimately appear as noise in the output of the laser-plasma accelerator.

As discussed in chapter 2, the surrogate model based approach of BO gives it a natural way to deal with measurement noise. The probabilistic Gaussian process model, can naturally express observation noise (see section 2.2.3), and specialized acquisition functions like *noisy expected improvement* [103] (see section 2.3.3) can take it into account when suggesting the next measurement. However, the presence of noise means an inevitable reduction of information per data point and could therefore also negatively impact the performance of BO.

Shot-to-shot variations of the laser parameters that are online monitored can be treated differently to simple observation noise. Although not avoidable to a certain extend, they can be measured for each event through the online diagnostics. This means that, for example, a change of the beam charge can be precisely traced back to a variation of the focus position. For the surrogate model of the optimizer this means, that there is a consistent mapping between the input parameter and the output. Consequently, the surrogate model does not need to introduce noise to explain the observed data. This is not the case for hidden parameters that are not covered by the online diagnostics as they would have an effect on the output that seemingly is not caused by changes to the input.

The challenge, however, lies in the fact that only the average state of the machine can be adjusted (e.g., mean laser energy and focus), which means that the optimizer may not get data for the exact machine state it requests. This discrepancy can particularly hinder the optimization's performance when aiming for a narrow optimum.

The precise influence of noise and jitter on the optimization process varies depending on the particular characteristics of the problem and is difficult to anticipate in general. To evaluate

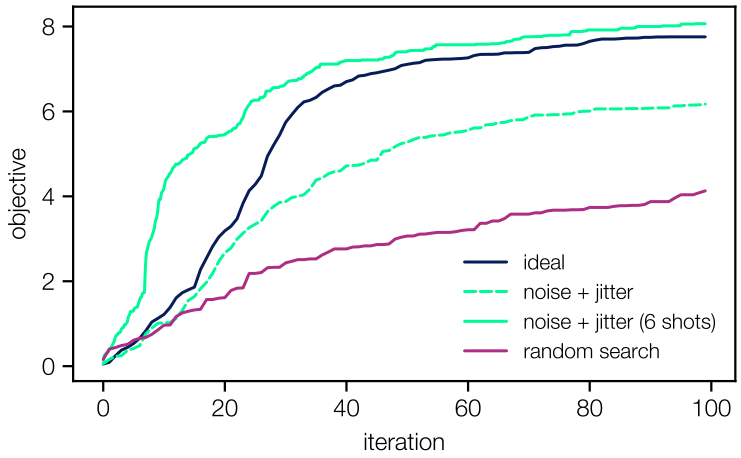

Figure 5.1 – Convergence tests for Bayesian optimization on LPA surrogate model (see section 3.2.4). Each line denotes the average performance (maximum objective found so far) over 50 independent optimization runs under each scenario. Experimental conditions (green lines) are simulated by introducing a fluctuation of 5 % on the laser energy and 150 µm on the focus position and by adding normally distributed noise ($\sigma = 0.1$) to the measured objective. The different lines indicate different optimization approaches and scenarios: Standard Bayesian optimization of the ideal (blue) and noisy (green dashed) setup, purely random selection of different setpoints (red), and Bayesian optimization of the noisy setup but collecting six measurements at each setpoint (green).

whether these effects significantly hamper performance in our experiment, the system is approximated using the surrogate model from the optimization of LPA PIC simulations in chapter 3. This surrogate model is presumed to serve as a reliable representation of the actual optimization problem in the experiment. It enables the modeling of the system under different scenarios, such as with or without noise, providing valuable insights into the impact of these factors. The performance of BO loops under different conditions is depicted in fig. 5.1. The unmodified predictions of the surrogate model act as a baseline comparison for the ideal stable and noiseless case.

To simulate the condition of the experiment, a shot-to-shot jitter of 5 % and 150 µm is added to the laser energy and focus position before querying the model, respectively. The objective value predicted by the surrogate model is obscured with normally distributed noise ($\sigma = 0.1$). Figure 5.1 shows the average performance over 50 independent runs for each scenario. The results of the individual runs can differ from each other due to the random noise and because the optimization is started by measuring six randomly selected setpoints.

The tests show that noise and jitter in fact have a negative impact on the performance of the optimization. The average objective found within 100 iterations (green dashed line) is 20 % smaller than without deteriorative effects (blue line). However, it is still significantly better than if the machine is tuned with a purely random search heuristic (red line).

Assuming that the performance of BO degrades due to the optimizer facing uncertainties from noise and not receiving the requested points, it would be logical to consider collecting more

data for each setpoint. This increased data collection would serve two purposes: reducing local uncertainty and utilizing the jitter to explore multiple points around the desired setpoint. This strategy in fact dramatically improves the performance as shown in fig. 5.1 (green solid line). By collecting six shots at each setpoint and using them to inform the surrogate model, the original performance of BO in the ideal case can be recovered and even surpassed. It should be noted that this comes at the cost of runtime, but in the experiment the system can anyways not be changed on a shot-to-shot basis. Therefore, it is reasonable to collect multiple events for each setpoint.

5.3 Beam quality optimization

Based on the surrogate model tests conducted in the previous section, it appears feasible to enhance the beam quality with Bayesian optimization in the experiment to a level and in a rate comparable to the simulations presented in chapter 3.

Mirroring the approach with the simulations, we used the objective $f = \sqrt{Q}\tilde{E}/\Delta E$ as a measure for the beam quality, specifically the spectral density. The exposed control parameters were the laser energy and longitudinal focus position as well as the three gas flows in the plasma source ($1 \times N_2$, $2 \times H_2$). These five parameters resembled the same degrees of freedom that were used in the simulations presented in chapter 3. Moreover, they allowed us to control the pressure difference between the front and back of the plasma source, a critical factor in determining the transition between mixed and pure gas regions. The permitted range for each of the parameters is given in table 5.1.

Using the online measured laser energy and defocus, the actual shot-to-shot machine inputs were mapped to the measured objective function. Thereby, as discussed in the previous section, the surrogate model could attribute fluctuations of the objective function to variations of the recorded laser parameters and the model could be trained with consistent single shot data. During the experiments the shot-to-shot rms jitters of the laser energy and focus position were 1.6 % and 200 µm, respectively. Noteworthy laser fluctuations, which were not directly considered during the optimization and therefore effectively treated as noise, included the pulse duration (2 % rms jitter), central wavelength (0.3 nm), and higher-order wavefront aberrations (<5 nm). Also not included in the surrogate model were the laser pointing and transverse focus position jitter on the order of 100 µrad and 5 µm, respectively.

Table 5.1 – Search space used for the Bayesian optimization.

laser energy	E_{laser}	$2.1 - 2.6$	J
focus position	Δz_{foc}	$-0.55 - 0.85$	mm
N_2 flow	$\dot{V}_{N_2}$	$0.0 - 3.0$	ml/s
H_2 flow front	$\dot{V}_{H_{2,\text{front}}}$	$0.0 - 3.6$	ml/s
H_2 flow back	$\dot{V}_{H_{2,\text{back}}}$	$0.0 - 6.0$	ml/s

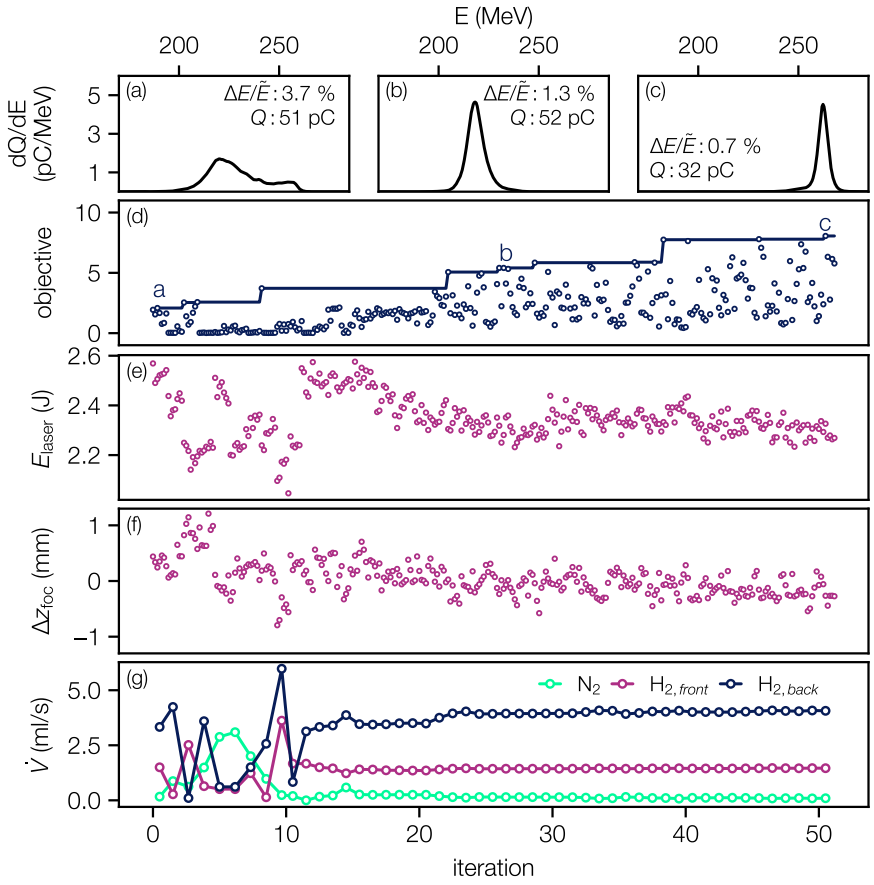

Figure 5.2 – Experimental optimization of LPA electron beams: (a-c) Measured energy spectra; (d) Measured objective function (dots) with the cumulative best result (blue line); (e-g) Input parameters with shot-to-shot measurements of the laser energy, focus position and gas flows for each input setting. The figure is adapted from [1].

In accordance to the optimal setpoint found in chapter 3, we set the quadrupole doublet to image 270 MeV electron beams onto the spectrometer. Beams with an energy far off this setpoint would be ignored by the optimization, due to the reduced energy resolution of the spectrometer and the larger measured energy spread (see section 4.5).

To start the optimization and initialize the surrogate model, we acquired data at 10 random input settings. For each iteration, the optimization gathered 6 shots and fed these individually to the surrogate model, maximized the acquisition function on the updated model, and then determined a new setpoint. Changes to the gas system took a few seconds to have an effect. Therefore, the machine was not tuned after every shot but by measuring 6 shots at each machine setting the jitter of the laser parameters could be used to locally explore the objective function (see section 5.2).

Figure 5.2 shows the evolution of the system during the optimization process. The initial beam, panel (a), has a spectrum with a small peak and a pedestal towards higher energies,

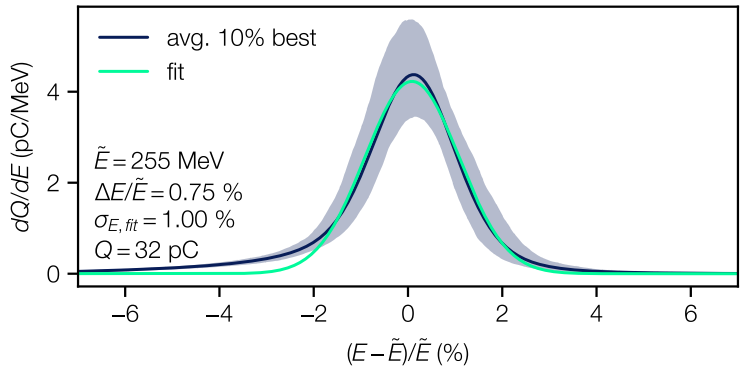

Figure 5.3 – Energy spectrum at optimized settings: averaged spectrum (black) and 5th-95th percentile (grey area) of the best 10 % of 2500 measured shots with a Gaussian fit ($\sigma_{\mathrm{E,fit}}$,green) and corresponding statistics. The energy spectra are normalized to their respective median energy, to eliminate the effects of shot-to-shot energy variations (1.5% rms) from the statistics. Over the considered spectra the average median energy is 255 MeV, the average energy spread is 0.75 % median absolute deviation or 1 % rms based on Gaussian fits. The average charge is 32 pC. The figure is adapted from [1].

which indicates that the beam still has a correlated energy spread from un-matched beam loading. After a few iterations, panel (b), a well-defined peaked spectrum is measured. Slight adjustments increase the energy of the beams to ∼250 MeV, panel (c). The decrease of the relative energy spread that comes at the cost of beam charge results in the improvement of the objective function. By balancing the relevant laser and plasma parameters the algorithm was able to find a setting with sub-percent energy spread starting from initially random conditions after a runtime of 45 min.

To thoroughly characterize the accelerator performance at this setpoint, we adapted the focusing to image the final beam energy of 250 MeV for best spectral resolution. Figure 5.4 shows the average spectrum of the best 10 % of the 2500 recorded events, which we consider representative for operation at the optimized settings. On average, these beams had 32 pC and an energy of 255 MeV with 1% rms (0.75% mad) energy spread. The average objective value of the selected shots was 7.6. Notably, these parameters closely align with the optimal state obtained in the simulations in section 3.2.3 ($\Delta E = 0.7$ %, $Q = 46$ pC) which confirms that BO performs well even under the unstable and noisy conditions of the experiment.

While beams with excellent quality are produced for a significant fraction of shots, a considerable shot-to-shot variation in the electron spectra is observed at the optimized setpoint, as demonstrated by the waterfall plot of sequential energy spectra in fig. 5.4a. A closer examination of the correlation between the beam energy, energy spread, and the charge, shown in fig. 5.4b-c, indicates that these fluctuations arise from variations in the beam loading conditions of the system. The correlations demonstrate an inverse dependency between energy and charge, along with a characteristic V-shape that was similarly observed in the model-based analysis of the simulated case in chapter 3. These correlations can be explained as follows: As more charge is injected into the wakefield, the average accelerating field along the bunch

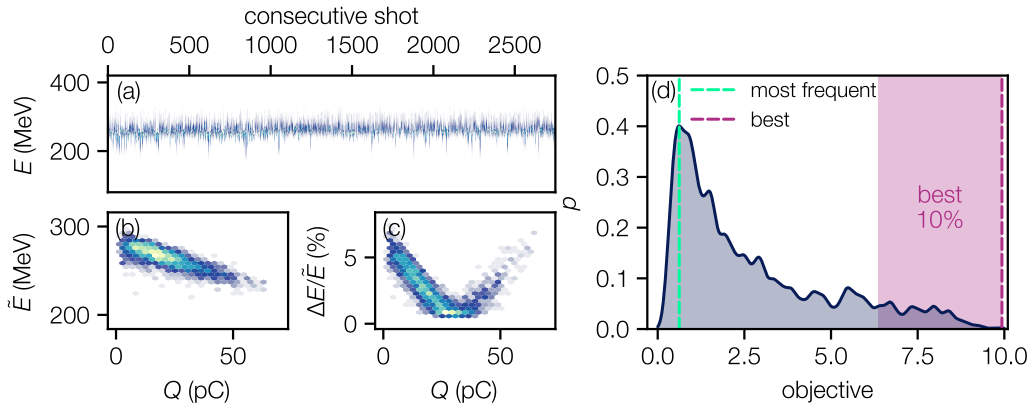

Figure 5.4 – Shot-to-Shot energy spectra at optimized setpoint. (a) Waterfall plot of consecutive electron energy spectra. (b) Correlation between beam charge and median energy. (c) Correlation between charge and energy spread (median absolute deviation). (d) Shot-to-shot distribution of the objective function. The red shaded area indicates the 10 % best beams used for fig. 5.3.

diminishes due to beam loading, leading to the observed energy reduction. The point at which beam loading negates the slope of the wakefield, resulting in an average flat accelerating field (refer to section 3.1.2), corresponds to the minimum of the charge and energy spread correlation, which in this case is at ∼30 pC.

The deviations from this point of optimal beam loading conditions are the result of the shot-to-shot fluctuations of the laser parameters [5]. By building the surrogate model on the basis of all individual events, the algorithm aimed at settings where the best possible beam could be generated, regardless of how frequently these would occur. Thereby, the optimizer fully ignores the distribution of the objective function, i.e., the stability of the setpoints. This can be seen when looking at the shot-to-shot distribution of the objective function at the optimized setpoint as shown in fig. 5.4d. Even though the setpoint can generate beams with high objective values, peaking as high as 10, the associated distribution is greatly skewed towards the lower end due to the inherent jitter of the laser. Consequently, the objective value most frequently encountered is notably low, only at 0.6.

5.4 Stability optimization

Demonstrating that we can identify and examine setpoints achieving the highest possible beam quality with the aid of Bayesian Optimization represents a significant achievement. These setpoints are particularly pertinent for investigating high-demand applications such as free-electron lasers and provide a glimpse into the future potential of LPAs as driver stability continues to improve. Yet, for most applications, the stability of the beam parameters holds equal importance to the achievable beam quality. Therefore, until the driver stability reaches the necessary levels, it is also favorable to find setpoints that are more robust against the fluctuations of current laser systems, even if this compromises the overall beam quality.

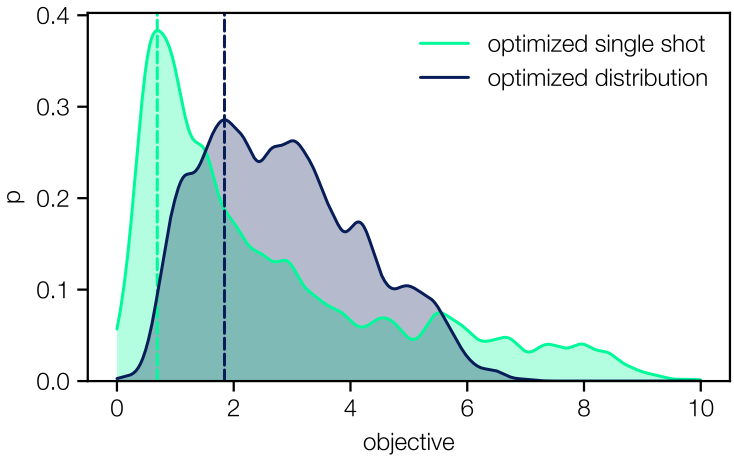

Figure 5.5 – Objective distributions of setpoints optimized with single shots (green) and with the mode of the distribution (blue). By optimizing the mode (dashed line) of the shot-to-shot distribution, the probability mass is shifted to higher objective values compared to the single shot method.

With the aim to discover if there are such setpoints that could improve the stability of the accelerator, we modified the optimization algorithm: instead of feeding the surrogate model with individual shots, we collected 20 shots at each setting. For each dataset we estimated the shot-to-shot distribution of the objective function, like the one shown in fig. 5.4d, through Gaussian kernel density estimation[1] [153]. From the given distributions, we identified the most frequent objective value, which was then used to train the surrogate model in conjunction with the averaged input parameters of the given setpoint. By optimizing this most frequent value, also known as the mode of the probability distribution, our aim was to shift the probability mass towards higher values, thereby reducing the occurrence of beams with extremely low objective values. The rationale behind these modifications is to incentivize the algorithm to enhance the overall distribution of electron beams, in the presence of the laser jitter, while dissuading it from identifying setpoints that exhibit high sensitivity to parameter fluctuations.

This approach differs from the one in section 3.3.1, where multi-objective optimization was used to balance stability and quality objectives, but essentially follows the same goal. Although this method provides less customization of the final outcome compared to the MOBO approach, it doesn't necessitate any modifications to the foundational optimization algorithm, making it straightforward to implement.

We ran the adapted BO method on the same search space (table 5.1) as the single shot optimization before. Figure 5.5 shows a comparison between the shot-to-shot distributions of the objectives of optimal setpoints. These setpoints were identified using single-shot optimization in the prior section, shown in green, and the modified method of optimizing the mode of the distribution, in blue. As intended, the peak of the distribution obtained by the modified

[1] Kernel density estimates are closely related to histograms. Instead of accumulating data into fixed bins, each data point is represented by a Gaussian function. The sum of these functions then gives the estimate for the probability distribution of the data.

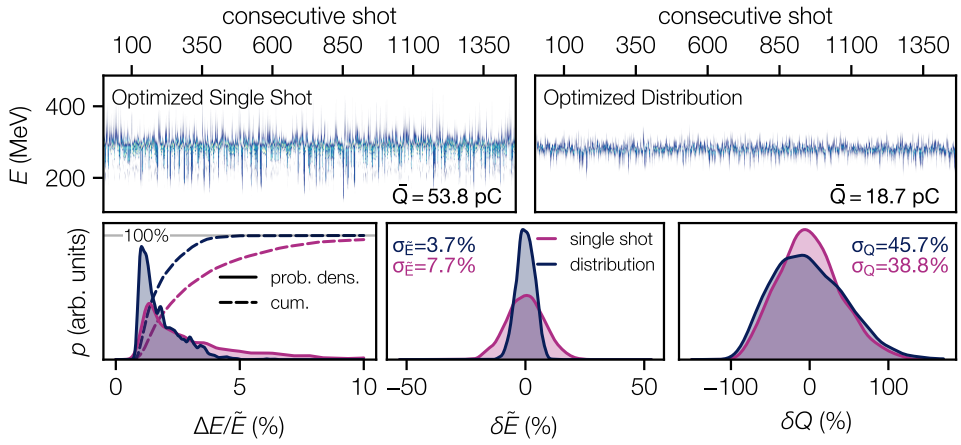

Figure 5.6 – Comparison between optimized setpoints. Top row: Waterfall plot of consecutive spectra for the setpoints optimized for highest quality (left) and for stability (right). Bottom row: Shot-to-shot distributions of the energy spread, energy and charge for the high quality (red) and stability (blue) setpoints. The distributions for the energy ($\delta\tilde{E}$) and charge (δQ) are the deviations from the respective mean values.

Table 5.2 – Input parameters for stable and reference setpoint and the resulting on-axis pressures in the plasma source.

		MODE	SINGLE SHOT
laser energy	E_{laser}	2.5 J	2.5 J
focus position	Δz_{foc}	0 mm	0.09 mm
N_2 flow	$\dot{V}_{N_2}$	0.083 ml/s	0.16 ml/s
H_2 flow front	$\dot{V}_{H_{2,\text{front}}}$	2.17 ml/s	1.5 ml/s
H_2 flow back	$\dot{V}_{H_{2,\text{back}}}$	3.83 ml/s	6.0 ml/s
pressure front	p_{front}	15.9 mbar	18.9 mbar
pressure back	p_{back}	18.2 mbar	25.5 mbar

optimization method is at larger values compared to the setpoint found in the previous section. Importantly, the probability for beams with small objective values falling below 1 is significantly reduced. However, also the largest possible objective value is decreased.

The distributions in fig. 5.5 confirm that the modified BO algorithm operates as we anticipated. The more relevant question, however, is how the changed distribution of objective values relates to the physical quantities of the electron beams. To this end, fig. 5.6 compares the relevant beam properties of the setpoint identified through the mode-optimization method to a reference setpoint similar to the one discussed in the previous section, which was manually fine-tuned to achieve a comparable beam energy. The input parameters for the two setpoints can be found in table 5.2. Both settings share similar laser conditions, but the stability-

optimized setting has a lower N_2 concentration (3.6% compared to 10%), and a reduced pressure in the plateau.

As clearly visible by the waterfall plots in fig. 5.6, the stability of the electron energy spectrum amidst the shot-to-shot fluctuations of the drive laser is markedly enhanced: 90 % of the accelerated beams maintain an energy spread smaller than 5 %, compared to the 60 % in our reference scenario. Simultaneously, the energy stability sees an improvement of more than a factor of 2, while sharing a comparable charge instability of 45 % and 35 %.

Echoing the findings from the simulations in chapter 3, this is another demonstration of the adaptability of our setup. That is, it can either produce beams with the highest spectral density or generate beams of adequate quality that exhibit remarkable robustness against instabilities of the drive laser.

5.5 Simulations

To unravel the mechanisms that lead to the distinct characteristics of the two setpoints shown in fig. 5.6, we conducted particle-in-cell simulations that accurately mimic both scenarios.

Establishing simulation parameters that precisely mirror the experimental conditions poses a significant challenge. Given the sensitivity of our setup, inevitable systematic measurement errors—particularly in parameters such as laser energy, focus position and size, pulse duration, and plasma density—can yield simulation outcomes that markedly deviate from the experimental results. To mitigate these disparities, we leveraged Bayesian optimization to subtly adjust these error-prone input parameters. The optimizer compared the correlations of the beam charge, energy and energy spread to the experimental results by varying the focus position in the simulation and minimized the difference. The discrepancy of the input parameters of the simulation and the experiment found this way were on the order of 10 %. A more comprehensive explanation of our method to accurately determine input parameters with minimal bias can be found in the supplemental material of [5].

The results of the simulations for the two setpoints are illustrated in fig. 5.7, alongside the data obtained from the experiment. As shown in fig. 5.7(a-f) the simulations accurately reproduce the observed behavior caused by changes of the focal position. As anticipated, the focus position significantly influences the properties of the resulting electron bunch for both configurations of the LPA.

A noticeable linear decrease in the beam charge is observed as the focus shifts downstream, a result of the diminished injection volume caused by the reduced intensity in the mixed gas region, as previously discussed in chapter 3. For the stable setpoint the absolute dependency between charge and focus position is reduced, from $-77\,\mathrm{pC\,mm^{-1}}$ to $-27\,\mathrm{pC\,mm^{-1}}$. However, the average charge also experiences a reduction, leading to an equivalent relative charge sensitivity for both setpoints at 145 % per 1 mm focus shift. This aligns with expectations as the change in charge is primarily driven by the variation of the ionization volume resulting from the focal shift. Given that both settings employ nearly identical laser parameters, they consequently exhibit the same relative charge sensitivity.

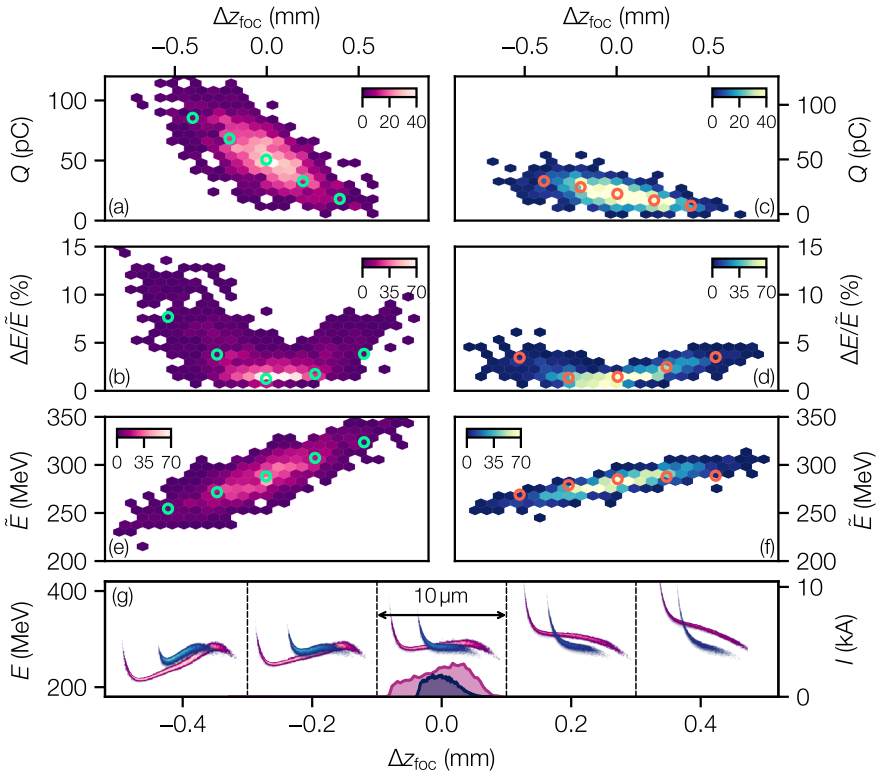

Figure 5.7 – Stability-optimized LPA: (a)-(f) Measured (heatmaps) and simulated (circles) correlations between the laser focus position and beam parameters for a typical reference (purple) and the stability optimized (blue) setting. Colormaps indicate the number of recorded events. (g) Longitudinal phase spaces from PIC simulations with the corresponding current profiles. The figure is adapted from [1].

However, when it comes to the spectral properties of the bunch, the previously observed pattern does not apply. Influenced by the effects of beam loading, these properties hinge on not only the absolute bunch charge but also significantly on the current profile. When observing the stable setpoint, the V-shaped behavior of the energy spread appears notably more shallow, which consequently leads to less severe beam deterioration when deviating from the optimal point. This tendency is also mirrored in the median energy of the beams. For the stable case, the focus variation sensitivity is 45 MeV mm^{-1}, less than half of the 95 MeV mm^{-1} observed in the reference case.

This particular behavior can be understood by examining the shapes of the longitudinal phase spaces of the beams generated by the two specific setpoints, as depicted in fig. 5.7g.

In both setpoints, changing beam loading conditions—induced by the alteration in charge—lead to a noticeable change of the energy correlation (chirp) along the bunch. An upstream, or negative shift in the focus, and consequently an increase in charge, causes the tail of the bunch to lose energy compared to the head. Conversely, energy is gained in instances of downstream shifts away from the optimum. Consequently, the median energy of the bunches

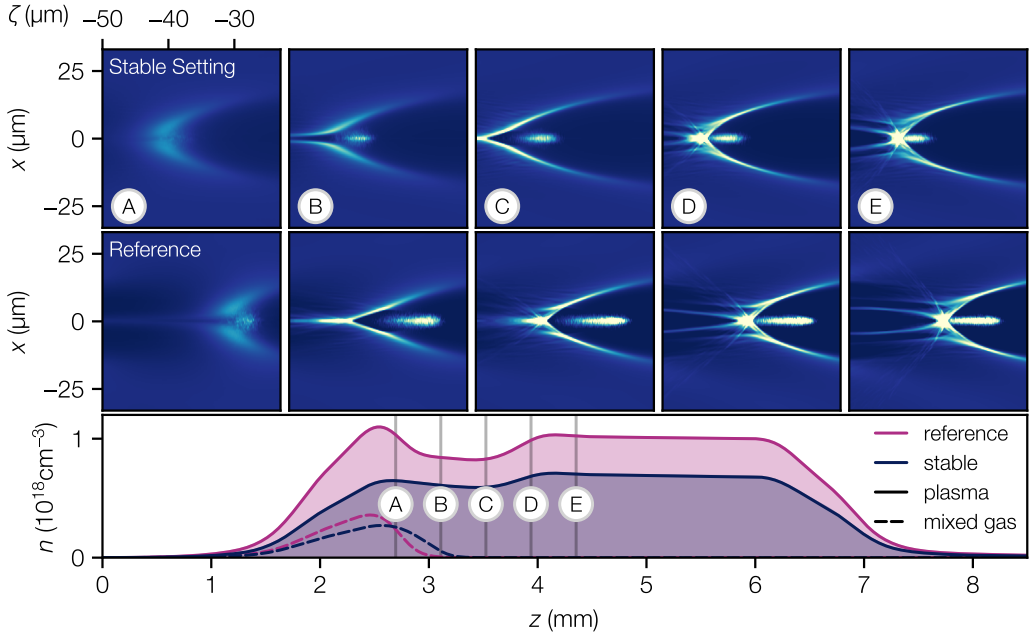

Figure 5.8 – Dynamics of the injection process. Top rows: Charge density (brighter is higher) of the wakefield at different points in the plasma source. Bottom: Plasma and mixed gas density for the stable and reference setpoints.

either diminishes or increases, and the projected energy spread deteriorates when departing from the optimum.

Interestingly, the chirp—defined as the change in energy per unit bunch length—remains virtually identical for corresponding cases of the stable and reference setpoints. The notable distinction lies in the fact that beams originating from the stable setpoint are significantly shorter, as evident in the comparison of current profiles in fig. 5.7g. This reduction in bunch length translates into a lesser impact on the projected energy spectrum when the phase space undergoes the shearing motion induced by mismatched beam loading.

The shortened bunch length can be traced back to the plasma density profile, a comparison of which is depicted in fig. 5.8. For the stable setpoint, which features a reduced N_2 concentration of 3.7 % in the mixed gas in contrast to the 10 % found in the reference case, the plasma density peak at the start of the plasma source—resulting from the nitrogen's lower ionization levels—is considerably less pronounced. Indeed, the plasma density in the mixed gas region is even lower than that in the subsequent plateau.

This difference in the plasma profile profoundly influences the injection process dynamics, as evidenced by the charge density plots in fig. 5.8. Owing to the reduced overall plasma density and the correspondingly longer plasma wavelength, the injection is initiated at a larger distance ζ behind the drive laser (frame A). As the mixed gas concentration decreases, thereby reducing the plasma density (frame B), the back of the wakefield falls back and with it the phase of the newly injected electrons. The injected bunch increases in length.

The non-linear relationship between the plasma wavelength λ_p and the density n_0, i.e., $\lambda_p \propto 1/\sqrt{n_0}$, along with the smaller overall density difference, results in a lesser phase advance in the lower density stable setting compared to the reference case. Consequently, the length of the injected bunch is shorter at the end of the mixed gas region (frame C). Following this, the plasma density rises again into the plateau region (frame D), causing the wakefield's peak to advance towards the injected bunch's back, and even truncate a portion of the bunch tail.

For the stable setpoint, where injection is triggered at a density lower than in the plateau, the bunch head lies in a phase that would normally be truncated by the wake's back when the wakefield contracts. The bunch only survives due to the convergent evolution of the drive laser, which in the non-linear regime increases the plasma wavelength. This culminates in the creation of a short bunch in the wakefield when entering the plateau region for acceleration (frame E), with the bunch for the stable setpoint being approximately half the length of that in the reference case.

5.6 Summary

In summary, we've successfully demonstrated the use of Bayesian optimization to tackle the intricate problem of identifying optimal operation points in a laser plasma accelerator. We discussed strategies to cope with the noisy and unstable conditions presented by the experiment. Running the optimization on single shot data collected at each setpoint allowed us to explore a regime of the setup that consistently provided electron beams with subpercent energy spread and significant beam charge.

By introducing a modified optimization objective that accounted for the shot-to-shot distribution of the system, we identified an operation regime which improved the shot-to-shot stability of the system by more than a factor of two. Through the correlations of the beam parameters and the corresponding particle-in-cell simulations, we linked this improved stability to a reduction in bunch length. This was achieved by meticulously fine-tuning the injection process, specifically by adjusting the plasma profile.

The data featured in this chapter originated from the initial experiments demonstrating the application of Bayesian optimization. Since then, Bayesian optimization has become a routine tool at LUX, employed regularly to reoptimize the system following larger changes to the plasma source or laser system. The setpoints found this way can typically be used for many beam times spanning over several weeks.

In conclusion, our application of Bayesian optimization for adjusting the parameters of a laser plasma accelerator has led to the discovery of two extremely beneficial operational regimes. The first regime delivers an exceptional beam quality, characterized by subpercent energy spread and high spectral density. The second regime significantly enhances the shot-to-shot stability, where 90 % of shots maintain an energy spread better than 5 %.

6 Tuning Curves for a Laser-Plasma Accelerator

This chapter presents the experimental demonstration of multi-objective Bayesian optimization (MOBO) employed to discover optimal tuning curves for a laser-plasma accelerator. These curves enable precise adjustments to system characteristics such as charge and energy, while simultaneously maintaining optimal energy spread. The findings discussed here have been published in *"S. Jalas et al., Tuning Curves for a Laser-Plasma Accelerator, Physical Review Accelerators and Beams (2023)* [2]. Similar to the previous chapter, this discussion is primarily a reproduction of that publication.

6.1 Overview

The previous chapter showcased how Bayesian optimization can serve as a powerful instrument for realizing exceptional performance from a laser-plasma accelerator. The selection of the objective function—whether it's based on single shots or the full distribution—permitted us to expose two very distinct characteristics of the setup. The strategy adopted in the previous chapter, though effective, lacks precision when forecasting the outcomes of the optimization, especially concerning beam parameters such as charge and energy. While the objective function can guide the general direction of desired beam characteristics, it leaves room for ambiguities. However, the ability to control these parameters with high precision and independence is indispensable for a variety of real-world applications that demand specific beam properties to meet their unique requirements.

LPAs have previously demonstrated some tunability [154] but reaching the desired level of control is a very challenging task: Multiple tuning parameters can act directly or indirectly on the same beam property, and beam parameters may influence or even counteract each other through the complex laser-plasma interaction. For example as shown in the previous chapters, beam loading [91, 47, 155, 5] couples bunch charge and energy spread. Simply varying one property (e.g., higher charger) can easily deteriorate (e.g., higher energy spread) the other.

Consequently, tuning the electron beam phase space represents a multi-objective optimization problem, with the goal to identify solutions that provide the best trade-offs between multiple, conflicting objectives. The set of all of these optimal trade-offs forms a hypersurface through the space of possible solutions that is commonly referred to as the Pareto front (see section 2.3.4). Gaining insight into the Pareto front is crucial for achieving optimal beam properties, as it helps to identify and reconcile the underlying conflicting mechanisms.

This chapter will present Pareto-optimal tuning curves to control the energy, charge, and energy spread in our laser-plasma accelerator. We utilize multi-objective Bayesian optimization [106, 110, 111] to sample the system's Pareto front and derive tuning curves that provide

accurate control over the relevant beam properties. We demonstrate the concept using simulations and present a method to handle the added complexity from noise under experimental conditions. Our experiments show that we can maintain optimal beam loading conditions over a broad range of beam configurations, allowing for flexible charge adjustments while minimizing negative impacts on energy spread at different specific design energies.

In section 6.2, a brief return to simulations is made to elucidate the concept and its potential. Following that, in section 6.3, the application of MOBO in the experiment will be discussed, along with a presentation of the experimental findings. Lastly, in section 6.4, a tuning model that enables precise adjustment of LUX's electron beam parameters is derived from the MOBO results.

6.2 Simulations

The application of MOBO to simulated LPA setups was already discussed in section 3.3. However, the discussion there focused on the generalization of the optimization procedure to take multiple beam parameters into account simultaneously. In practice, it can be useful to quickly switch between optimal states of the setup and importantly to seamlessly tune between them. The following discussion will focus on this aspect.

To do this, the LUX setup is simulated once more. This time the setup employs the revised version of the plasma source (as referred to in section 4.4) and incorporates argon as the mixed gas dopant.

The goal is to begin from a point of optimal beam loading conditions with the minimum energy spread, and then adjust the beam charge while maintaining a low energy spread. As previously elaborated, an uncomplicated method to modify the bunch charge in the given setup is to alter the focal plane of the drive laser: Typically, the focus is set to the second half of the plasma profile to maximize the wakefield strength in the acceleration region. Shifting the focus upstream towards the mixed gas region increases the injection volume (ionization of inner shell Ar electrons) and thus beam charge. This effect is shown in fig. 6.1a, continuously shifting the focus upstream over a range of 300 μm. Starting off with a beam that is optimally beam loaded with 50 pC at 0.7 % energy spread, the charge increases as intended. However, with this simple increase in charge, the previously flattened wakefield becomes overloaded and deforms, causing the bunch tail to experience a lower accelerating gradient than the head. As a result, the energy spread rapidly grows to unacceptable levels.

In principle, it is possible to counter-act this effect by increasing the wakefield strength. However, parameters that control the wakefield do not act in an isolated manner: For instance, raising the plasma density in return affects self-focusing and the evolution of the drive laser. Further adjustments of the laser energy and focus position would be required to recover the injection conditions; yet, they also influence the shape of the accelerating wakefield, which in turn could make a different bunch current profile necessary to achieve optimal beam loading. This requires adjustments of the Ar-concentration, which not only changes the injection rate, but also the current profile by affecting the plasma profile in the mixed gas region.

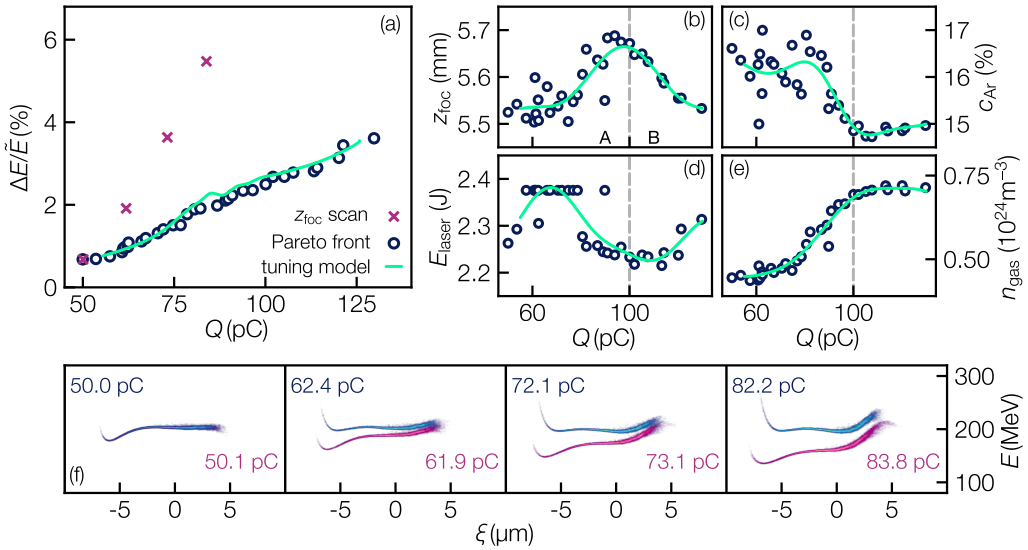

Figure 6.1 – LPA Tuning Curves (Simulations): The LPA setup is optimized for both charge and energy spread at 200 ± 10 MeV using MOBO. Panel (a): energy spread, tuning the charge with the laser focus position only (crosses), or by a combined complex tuning of many parameters (circles), which are individually shown in panels (b)-(e). (A) and (B) mark two different tuning strategies. Panel (f): longitudinal phase spaces for the optimized tuning (blue) and the focus scan (violet). Solid lines, panels (a)-(e) mark the continuous tuning curves, derived from the optimization. The figure is adapted from [2].

This cascade of consequences illustrates, that manually searching the functional dependencies that result in tuning of the beam charge while maintaining optimal beam loading and energy spread is futile. Instead, we apply MOBO to systematically study the input parameters that make up this optimal tuning behavior. Here, we chose beam charge Q, relative energy spread $\Delta E/\tilde{E}$ (median absolute deviation), and the deviation from a design energy of 200 MeV, as objectives for optimization.

The results of this optimization, consisting of 700 individual simulations, are shown in fig. 6.1a. Compared to the simple focus scan, we can increase the charge while keeping the energy spread low. As expected, to achieve this Pareto-optimal tuning, the input parameters (laser energy, focus position, gas density, and Ar concentration) need to be varied in non-obvious ways as shown in fig. 6.1b-e.

Notably, for lower charges (region A) the focus position z_{foc} and Ar concentration c_{Ar} have to be tuned in opposite directions from what would be expected when increasing the charge with these parameters individually. Instead, the focus is moved away from the mixed gas region, the Ar concentration is reduced, and gas density and laser energy are increased. The higher gas density increases the plasma density and the amount of Ar. Together with the higher laser energy, this results in an increase of the injected charge and a modification of the wakefield strength, which prevents excessive beam loading. Shifting the focus downstream prevents diffraction of the drive laser towards the end of the plasma and maintains optimal beam

loading throughout the acceleration process. Moreover, a slight decrease in Ar concentration is necessary to match the injected current profile to the wakefield at the increased plasma density.

The tuning behavior changes completely once the gas density approaches the upper limit of the tuning range ($0.72 \times 10^{24} \mathrm{cm}^{-3}$) at around 100 pC (region B). Now, to even further increase the charge, the focus needs to be tuned in the opposite direction, i.e., upstream, and the Argon concentration needs to be increased again slightly. The laser energy that was previously also reaching the maximum of the permitted range is slightly reduced.

As shown in fig. 6.1f, this way of precisely balancing the input parameters maintains a flattened phase space over a wide range of charges, while keeping the beam energy fixed. The residual energy spread increase is primarily caused by correlated energy spread, building up only in the head and tail of the bunch. In contrast, the beams from the simple focus scan develop a significant energy correlation and energy deviation.

To derive an actual tuning curve that can be used in practice, the discrete, optimized working points have to be generalized. We fitted a multivariate kernel ridge regression model [1] [156] to the data, that maps the desired beam charge Q to the set of input parameters $\mathbf{X}$, i.e., laser energy, focus position, gas density, and Ar concentration, that provide this charge at Pareto-optimal energy spread: $\mathbf{X}_{\mathrm{pareto}}(Q)$. The model is shown in fig. 6.1b-e, with the solid line indicating the continuous variation in input parameters as we continuously increase the beam charge. Using these tuning curves, the beam energy spread closely follows the Pareto front, compare fig. 6.1a (solid line). Following our tuning curve, we can precisely set the beam charge, while providing the best possible energy spread and retaining the design energy.

6.3 Multi-objective Bayesian optimization at LUX

Realizing the same concept experimentally is significantly more challenging, due to added measurement noise and shot-to-shot variations (jitter) in the experimental conditions.

Conceptually, several sources of noise need to be distinguished, as discussed in the previous chapter. First, the optimizer might set the machine to a certain target state (e.g, laser energy and focus position), that could be different for the actual shot. We can account for this with online shot-to-shot measurements of the laser parameters. Second, all measurements suffer from at least some measurement noise, that we cannot avoid. This random noise leads to uncertainty in the predictions of the surrogate model that can only be mitigated with additional data.

To navigate the shot-to-shot fluctuations and inherent noise, data from 40 shots were recorded at each setpoint. Merely taking the average from these shots would inherently consider the shot-to-shot laser fluctuations as random noise. This would result in the introduction of

[1] Kernel ridge regression is a method in machine learning that introduces a kernel function to perform regression in a higher dimensional feature space. This essentially allows the algorithm to model complex, nonlinear relationships between variables. It's comparable to Gaussian process regression but provides a single function instead of a distribution over functions.

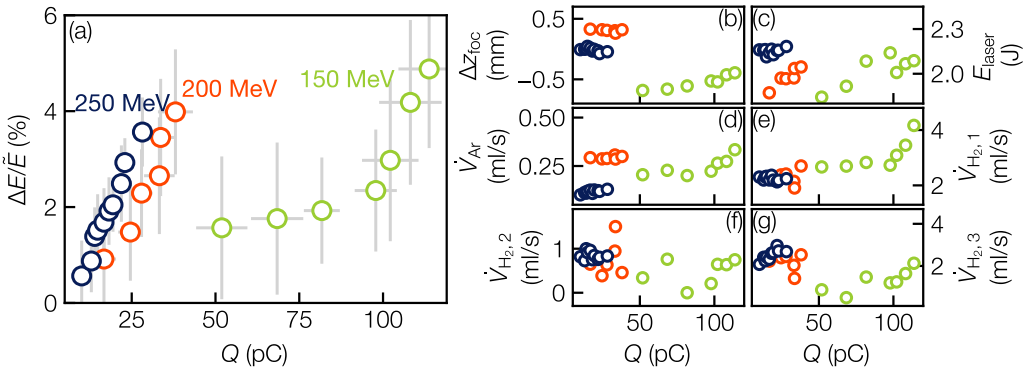

Figure 6.2 – Experimental results: (a) Pareto fronts of the beam charge and energy spread for beams within ±10 MeV of the target energies with corresponding input parameters (b-g), i.e. the shift of the focus position Δz_{foc}, laser energy E_{laser} and the flows of the gas supplies $\dot{V}$. The figure is adapted from [2].

extra uncertainty into the surrogate model, an outcome that isn't desirable. Instead, at each working point, we trained a local Gaussian process model using the online measurements of the laser energy and focus position as inputs and the electron parameters as outputs. This allowed us to disentangle the effects of the laser jitter from the remaining noise and interpolate the properties of the electron beam for the exact input parameters that were requested by the optimizer. Then we used this information to train the surrogate model that guided the optimization.

The employed method shares similarities with the single-shot approach discussed in the previous chapter in that it describes the accelerator at a singular point in the parameter space. However, due to the computational expense associated with using the MOBO algorithm on extensive datasets, this approach, using local modeling, was chosen. An alternative would have been to use the mode-based method from before, potentially constraining the optimization results to more stable setpoints. However, as our goal was to achieve a more unfiltered depiction of the accelerator's performance, this alternative was disregarded.

To explore the capabilities of our setup, we performed a series of experiments with MOBO runs at 150, 200 and 250 MeV. As before, the objectives for the optimization were the beam charge, relative energy spread, and deviation from the reference energy. The resulting Pareto fronts with their corresponding input parameters after 100 iterations, i.e., 4000 individual shots, are shown in fig. 6.2.

For all three target energies we find low energy spread setpoints over a significant range of beam charges. Consistent with the simulations, the energy spread grows with increasing charge. For lower target energies, the achievable charge increases as one would expect from an aspect of energy conservation. As a result, for 150 MeV, it was possible to tune the beam charge between 50 and 130 pC, while maintaining a relative energy spread of less than 5 %. Up to a charge of 100 pC, the energy spread only shows a modest growth and stays below 2.5 % but rapidly increases from there on. This behavior implies that the available input parameter range did not support optimal beam loading beyond this point. Therefore, additional charge

can only be loaded into the wakefield at the cost of significantly distorting the accelerating field. For the two higher energy cases, this quick increase of the energy spread is present already at low charge, indicating that here beam loading conditions immediately worsen so much with increasing beam charge that a significant energy correlation is imprinted.

Figure 6.2b-g shows, that the three different energies populate very different islands in the input parameter space, especially for Δz_{foc}, $\dot{V}_{\mathrm{Ar}}$ and $\dot{V}_{\mathrm{H_2,3}}$. Also the way some parameters are tuned varies from energy to energy. For example, for 150 MeV, the focus is moved downstream to increase the charge similar to the lower charged beams in the simulated setup before. For the other two cases it is moved upstream towards the mixed gas comparable to the higher charge cases in the simulations. This again hints that these two cases are already in a regime that is dominated by strong beam loading. Assuming this is true, we expect that these cases would behave similar to the 150 MeV case when provided with more laser energy and exhibit lower energy spread over a wider tuning range. The fact that these three cases behave so diverse in regard to the input parameters highlights the complexity of, first, finding beams with acceptable quality and then, second, being able to control them.

A general feature that is shared by nearly all configurations in fig. 6.2 is that the flow through the last gas inlet $\dot{V}_{\mathrm{H_2,3}}$ is higher than through the middle one $\dot{V}_{\mathrm{H_2,2}}$. By creating a slight density ramp the slippage between the electrons and the laser can be compensated through contraction of the wakefield which is a way to increase efficiency and load more charge.

6.4 Tuning model

The Pareto fronts observed in our experiment, depicted in fig. 6.2, possess fewer sample points relative to the prior simulations discussed in section 6.2. This outcome stems from fewer iterations run by the optimizer, possibly further impacted by the more challenging conditions encountered during the experiment. Consequently, providing a continuous tuning curve is even more important. The data shows a less non-linear behavior than the simulations before, possibly because it is less converged to the true Pareto front of the system. In this case, it is sufficient to use a simple multivariate linear model to describe the main parameter variations.

We derived such a model, exemplary for 250 MeV,

$$\begin{aligned}\mathbf{X}_{\mathrm{pareto}}(Q) &= \left(\Delta z_{\mathrm{foc}}(Q), \dot{V}_{\mathrm{Ar}}(Q), \dot{V}_{\mathrm{H_{2,3}}}(Q)\right) \\ &\approx \left([59 - 4Q]\,\mu\mathrm{m}, \left[\frac{9.5}{10^2} + \frac{1.25}{10^3}Q\right]\mathrm{ml\,s^{-1}}, \left[1.8 + \frac{4}{10^2}Q\right]\mathrm{ml\,s^{-1}}\right),\end{aligned}$$

with Q in units of pC. Here we only considered parameters that made significant contributions to the tuning of the beam charge, keeping the remaining inputs constant ($E_{\mathrm{laser}} = 2.15\,\mathrm{J}$, $\dot{V}_{\mathrm{H_2,1}} = 2.24\,\mathrm{ml/s}$, $\dot{V}_{\mathrm{H_2,2}} = 0.84\,\mathrm{ml/s}$). For example, to generate beams with 10 pC, according to the tuning curve, an operator would set the machine to $\Delta z_{\mathrm{foc}} = 19\,\mu\mathrm{m}$, $\dot{V}_{\mathrm{Ar}} = 0.107\,\mathrm{ml/s}$ and $\dot{V}_{\mathrm{H_{2,3}}} = 2.2\,\mathrm{ml/s}$.

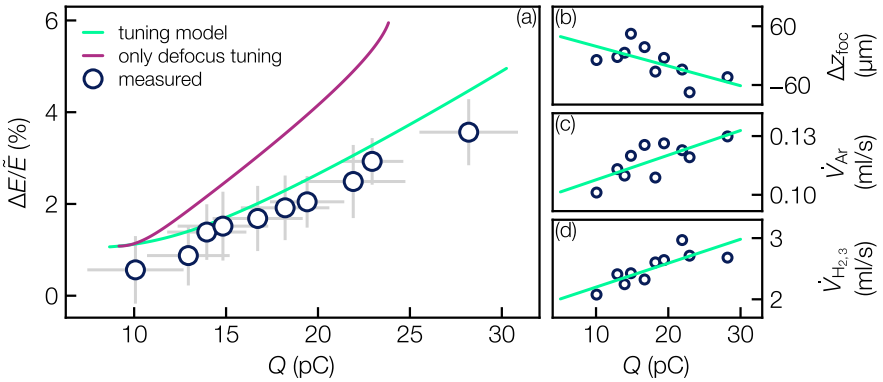

Figure 6.3 – Beam charge tuning curve at 250 MeV. A linear model (green line) is fitted to the optimization result mapping the requested beam charge to the corresponding input parameters along the Pareto front (b-d). (a) The tuning curve was validated and compared to the measured data (circles) with predictions from a Gaussian process surrogate model. For comparison the effect of a laser focus position scan over 450 µm is shown (red line). The figure is adapted from [2].

The input parameters over the full range of the tuning curve and the underlying data is shown in fig. 6.3b-d. The tuning strategy is to shift the focus towards the mixed gas region and to increase the flow of argon to enhance the beam charge. To support the additional charge and balance beam loading, the flow of hydrogen in the back of the plasma source needs to go up. This leads to a higher gas pressure in the plateau region and creates a density ramp towards the back of the plasma profile, ensuring optimal performance.

To evaluate the effectiveness of the tuning curve, we employ the surrogate model obtained from the optimization process to forecast the response of the LPA when adjusted based on the curve. For the example above, the surrogate model predicts that the LPA generates beams with 10.9 pC at 1.2 % energy spread that are within 4.3 MeV of the 250 MeV design energy.

The green line in fig. 6.3a shows the predicted behavior of the energy spread when scanning the charge according to the tuning curve. The resulting working points closely align with those measured during optimization, indicating their proximity to the presumed Pareto front of the system. Remarkably, this behavior is achieved using only the three most important input parameters identified through our analysis. In contrast, when tuning the charge only with the focus position, the resulting growth of energy spread is roughly a factor of two larger than what is obtained with our tuning curve. Provided with the simple tuning curve, operators can precisely set the beam charge required for their application, while still achieving the smallest possible energy spread.

6.5 Summary

In summary, we have demonstrated dedicated control over electron bunch parameters in our laser-plasma accelerator using continuous tuning curves, both in simulations and experiments,

that have been build via multi-objective Bayesian optimization of the machine. In particular, the bunch charge was varied over a range of nearly 100 pC and at different electron energies (150-250 MeV), while providing the smallest possible energy spreads (<5 %) by moving the laser and plasma control parameters along Pareto-optimal tuning curves. Even more importantly, we could show that the derived tuning curves itself are meaningful expressions of the control and output parameter relationships, which allowed us to identify and explain how different physics mechanisms have to be balanced to provide optimal trade-offs between competing beam parameters.

Conclusion

If laser-plasma accelerators live up to their promise—providing electron beams of a quality that rivals conventional accelerators, but in a package compact enough for a small lab or portable enough to be transported in a truck—the implications for society will unquestionably be profound.

Plasma accelerators have made significant strides in recent proof-of-concept experiments, demonstrating an array of promising features. However, the robustness and reliability they have exhibited still falls short of the established performance standards set by their conventional counterparts. For laser-plasma accelerators to truly advance and gain widespread adoption, they need to consistently deliver the precise parameters that users demand. This isn't about meeting performance criteria occasionally, but consistently producing results with every single shot on every single day.

As the affordability of these machines increases, their accessibility will likewise expand, leading to an inevitable diversification in user demands. This evolution means that laser-plasma accelerators need to offer not only high-fidelity performance, but also flexible tuning capabilities. The capacity to adjust and vary beam parameters to cater to a multitude of applications is a crucial attribute that will heavily impact their broad utility and ultimate societal benefit.

Undoubtedly, while advancements in technology, such as more stable driver technology operating at kilohertz rates, will likely alleviate some of the challenges, they alone are not a universal solution. The complexity of laser-plasma accelerators is expected to increase, not decrease, potentially complicating their operation further. This trend underscores the necessity for systematic strategies to not only operate but also fine-tune these machines. Mere technical progress can improve certain aspects, but the demand for precise, reliable, and flexible control over beam parameters calls for a dedicated approach. This need is where the methodology introduced in this thesis, specifically the application of machine learning techniques, and more precisely Bayesian optimization, comes into play.

Bayesian optimization, with its unique strengths such as its robustness when dealing with sparse and noisy data, emerges as an ideal tool for navigating the intricate parameter spaces that characterize laser-plasma accelerators. As discussed in chapter 3, it can play a pivotal role in transforming particle-in-cell simulations into versatile design instruments. This technique allows for quick optimization of the degrees of freedom in a proposed setup, providing a fair assessment of the potential of any given idea. In this particular case, an LPA setup was optimized that utilized localized ionization injection for the generation of the electron beams. Through the optimization it was possible to quickly find parameters that in the simulation generate beams with 0.7 % energy spread and almost 50 pC charge, resulting in 2.5 kA beam current and a spectral density of 7 pC MeV^{-1}.

While it's important to note that other forms of optimization could theoretically achieve similar results—albeit at higher cost—Bayesian optimization stands out due to its surrogate model. This model offers more than just an optimized result—it provides insightful and interpretable data about the system being optimized. This capability to reveal the underlying mechanisms and relationships of the parameters within the system is a distinctive advantage of Bayesian optimization. Thus, it provides not only the means to refine and improve designs, but also deepens our understanding of the systems themselves. This concept was demonstrated in chapter 3, where the surrogate model was employed to understand the effects and interplay of various input parameters, and in particular to study the sensitivity of the optimized LPA setup to variations of the laser and estimate its stability under realistic conditions.

The model-based analysis revealed the system's significant sensitivity to changes in the longitudinal focus position of the drive laser. Such shifts can cause variations in the beam charge, which in turn impacts the beam loading conditions. Consequently, the spectral properties of the beam can change.

Through the application of multi-objective optimization it was demonstrated that the same system can support operational modes that are far less sensitive to such fluctuations in the focus position. Notably, it was shown that the overall quality of the electron beams can be balanced against shot-to-shot stability, effectively allowing for tailoring to match an application's specific tolerances. This insight showcases the flexibility inherent in the system when optimized effectively. Further, the use of multi-objective optimization and the Pareto front has proven to be a reliable method for comparing different design concepts. By contrasting the impacts of different dopant gases on beam stability and quality, the advantages and disadvantages of each choice were clearly delineated.

The latter portion of this thesis involved the practical application of the Bayesian optimization approach to a real laser-plasma accelerator at LUX. Following a concise introduction to the experimental setup of the LUX accelerator in chapter 4, chapter 5 demonstrated the autonomous tuning of the machine. This autonomous tuning successfully achieved a consistent setpoint, resulting in electron beams with subpercent energy spreads and approximately 30 pC of charge. The resulting spectral densities reached up to $5\,\mathrm{pC\,MeV^{-1}}$—beam parameters remarkably close to those achieved in the idealized simulations before. Mirroring observations from the simulations, the beams also showed a strong sensitivity to changes in the laser conditions.

Adjusting the objective function to account for the full distribution of the produced electron beams in light of the laser jitter, a stable regime like those seen in simulations was attained experimentally. In this stable regime, an impressive 90 % of the shots achieved an energy spread of less than 5 % and the shot-to-shot energy stability was 3.7 %. Detailed particle-in-cell simulations and analysis of the correlations between beam parameters revealed the source of this enhanced stability. It was found to be due to a reduced bunch length achieved through careful adjustment of the plasma profile.

With the achievement of improved stability and beam quality, in chapter 6 this thesis finally addressed the challenge of precisely tuning the beam parameters of the LPA. Multi-objective Bayesian optimization was introduced as an effective tool for this purpose. It enabled the identification of tuning curves for the accelerator that modified beam parameters, such as

bunch charge, in accordance with the system's Pareto front. This approach ensured optimal energy spread while tuning the desired beam parameters.

The versatility of this technique was demonstrated by tuning the electron beam charge over a broad range—nearly 100 pC—while maintaining optimal beam loading conditions and keeping the energy spreads below 5 % for energies between 150 and 250 MeV. The derived tuning curves were valuable in elucidating the sometimes counter-intuitive relationships between the control variables of the laser and plasma.

The methods developed in this thesis have since been adopted as standard tools for the operation of LUX. Whether it's modifications to the plasma source, adjustments to the laser, or even adaptations for new experiments in different energy regimes, Bayesian optimization has proven valuable in restoring previous beam parameters and swiftly adjusting the machine to suit the demands of the experiment at hand.

The broader impact of the research carried out in this thesis is evident already now, shortly after the first demonstration of Bayesian optimization in laser-plasma accelerators [129, 1]. The method has gained substantial interest within the community, evidenced by numerous publications that have leveraged this approach [6–8, 157–161]. Notably, simulations for a potential plasma based injector of the planned PETRA IV upgrade at DESY are being based on this methodology [6, 162].

Meanwhile, Bayesian optimization and machine learning-based strategies have been identified as important instruments in advancing the field and enabling potential future plasma-based facilities. This perspective is echoed in both the European Strategy for Particle Physics [163] and multiple contributions to the Snowmass 2021 Accelerator Frontier in the US [164–166], further emphasizing the recognition and embrace of these methods.

The imminent era of high repetition rate drive lasers, operating in the kHz range, undoubtedly signifies a transformative phase for the laser-plasma community. This development creates an opportune moment to proof the practical utility of laser-plasma accelerators. However, the benefits of a higher repetition rate can only be fully realized by also advancing our methods to keep pace with the increased data output. As data rates surge, it becomes vital to develop innovative techniques capable of managing the higher data throughput effectively. This challenge necessitates strategies that not only cope with the volume of data but also leverage this wealth of information to extract greater insight and drive further improvements.

Bayesian optimization, given its resilience in dealing with sparse and noisy data, will undoubtedly continue to be a key instrument in the toolbox of laser-plasma accelerator research. However, the higher data rates enabled by kHz repetition lasers will broaden the horizon of applicable machine learning techniques. With the luxury of more substantial datasets, methods that are more data-intensive can come into play. This includes reinforcement learning [167, 168], where machines learn to make specific decisions for example to act on environmental changes, or stochastic gradient descent [169], a potent optimization algorithm routinely employed in large-scale machine learning tasks, both of which could be adapted for real-time

accelerator tuning. Similarly, deep learning methods, known for their appetite for vast quantities of data, could become viable for uncovering complex, high-dimensional relationships in accelerator performance data [5].

As the field progresses, the repertoire of data analysis tools is expected to grow, offering a range of methods tailored to the particular challenges and opportunities of laser-plasma acceleration research.

The challenges and solutions being investigated for laser-plasma acceleration in this regard now reflect those also encountered in the realm of conventional accelerators [101, 170–174]. This parallel development signifies a departure from the established pattern where developments from conventional accelerators were predominantly imported into the field of laser-plasma accelerators, but rarely vice versa. Now, there exists a novel, bilateral opportunity for knowledge exchange between these two domains. This fresh perspective underscores a significant stride in the evolving synergy between the laser-plasma and traditional accelerator communities, paving the way for more robust and sophisticated technological advancement.

Acknowledgement

I'd like to extend my sincere gratitude to Andreas Maier and Wim Leemans for their guidance and supervision during my academic journey.

I joined Andreas Maier's group almost ten years ago, where I found an environment that not only encouraged the development of unconventional ideas but also provided the necessary support and direction to turn them into successful results. Andi's unshakable commitment to the well-being and interests of his team members has been a driving force behind our collective achievements and is something I hold immense admiration for.

The arrival of Wim Leemans at DESY marked a transformative moment for our campus, elevating it into a hub for plasma-based accelerators. I am profoundly grateful for his visionary leadership and the significant impact it has had on our work.

I owe a tremendous debt of gratitude to Manuel Kirchen, who served as my mentor when I first joined the group as a bachelor student and eventually became one of my closest friends. Our distinctive style of sharing and developing ideas, as well as our close connection, often drew good-natured ribbing from others over the years. Nonetheless, it's also the primary reason why I have developed such a deep passion for our work. Thank you!

Similarly, working alongside Philipp Messner, Paul Winkler, and Vincent Leroux, I've formed friendships that go beyond work. Having the privilege of working with such close friends has made those late-night control room sessions, sometimes stretching until dawn, something I genuinely look forward to. These bonds have not only enriched my life but have also made our shared experiences, both in our professional and personal lives, all the more meaningful. I deeply value your presence and the moments we've shared together.

I would also like to extend my sincere appreciation to all the current and past members of the former LUX, now MLS team. Reflecting on the years I've spent as part of this team, I'm truly amazed by the remarkable accomplishments of these individuals. Beyond their professional achievements, they are simply a fantastic group of people to be around.

To my parents and brothers, I want to express my heartfelt thanks for always being there for me. My late father, in particular, played a pivotal role in shaping my love for science and curiosity about the world. Your support and inspiration have been instrumental in making me who I am today.

To Jenny and Greta, words cannot adequately convey my gratitude for your constant love, unwavering support, and boundless understanding. Your presence in my life fills it with joy, purpose, and love. Thank you for being the pillars of my happiness.

Bibliography

[1] S. Jalas, M. Kirchen, P. Messner, P. Winkler, L. Hübner, J. Dirkwinkel, M. Schnepp, R. Lehe, and A. R. Maier, Bayesian optimization of a laser-plasma accelerator, Phys. Rev. Lett. **126**, 104801 (2021).

[2] S. Jalas, M. Kirchen, C. Braun, T. Eichner, J. B. Gonzalez, L. Hübner, T. Hülsenbusch, P. Messner, G. Palmer, M. Schnepp, C. Werle, P. Winkler, W. P. Leemans, and A. R. Maier, Tuning curves for a laser-plasma accelerator, Phys. Rev. Accel. Beams **26**, 071302 (2023).

[3] S. Jalas and A. R. Maier, Plasmasurfen auf pilotanlagenkurs, Physik in unserer Zeit **53**, 168–174 (2022).

[4] A. R. Maier, N. Delbos, T. Eichner, L. Hübner, S. Jalas, L. Jeppe, S. W. Jolly, M. Kirchen, V. Leroux, P. Messner, M. Schnepp, M. Trunk, P. A. Walker, C. Werle, and P. Winkler, Decoding sources of energy variability in a laser-plasma accelerator, Phys. Rev. X **10**, 031039 (2020).

[5] M. Kirchen, S. Jalas, P. Messner, P. Winkler, T. Eichner, L. Hübner, T. Hülsenbusch, L. Jeppe, T. Parikh, M. Schnepp, and A. R. Maier, Optimal beam loading in a laser-plasma accelerator, Phys. Rev. Lett. **126**, 174801 (2021).

[6] S. A. Antipov, A. F. Pousa, I. Agapov, R. Brinkmann, A. R. Maier, S. Jalas, L. Jeppe, M. Kirchen, W. P. Leemans, A. M. de la Ossa, J. Osterhoff, M. Thévenet, and P. Winkler, Design of a prototype laser-plasma injector for an electron synchrotron, Phys. Rev. Accel. Beams **24**, 111301 (2021).

[7] A. Ferran Pousa, I. Agapov, S. A. Antipov, R. W. Assmann, R. Brinkmann, S. Jalas, M. Kirchen, W. P. Leemans, A. R. Maier, A. Martinez de la Ossa, J. Osterhoff, and M. Thévenet, Energy compression and stabilization of laser-plasma accelerators, Phys. Rev. Lett. **129**, 094801 (2022).

[8] A. Ferran Pousa, S. Jalas, M. Kirchen, A. Martinez de la Ossa, M. Thévenet, S. Hudson, J. Larson, A. Huebl, J.-L. Vay, and R. Lehe, Bayesian optimization of laser-plasma accelerators assisted by reduced physical models, Phys. Rev. Accel. Beams **26**, 084601 (2023).

[9] D. C. Haynes, M. Wurzer, A. Schletter, A. Al-Haddad, C. Blaga, C. Bostedt, J. Bozek, H. Bromberger, M. Bucher, A. Camper, *et al.*, Clocking auger electrons, Nature Physics **17**, 512–518 (2021).

[10] M. Kirchen, R. Lehe, S. Jalas, O. Shapoval, J.-L. Vay, and A. R. Maier, Scalable spectral solver in Galilean coordinates for eliminating the numerical Cherenkov instability in particle-in-cell simulations of streaming plasmas, Phys. Rev. E **102**, 013202 (2020).

[11] V. Leroux, S. W. Jolly, M. Schnepp, T. Eichner, S. Jalas, M. Kirchen, P. Messner, C. Werle, P. Winkler, and A. R. Maier, Wavefront degradation of a 200 TW laser from heat-induced deformation of in-vacuum compressor gratings, Opt. Express **26**, 13061 (2018).

[12] N. Delbos, C. Werle, I. Dornmair, T. Eichner, L. Hübner, S. Jalas, S. W. Jolly, M. Kirchen, V. Leroux, P. Messner, M. Schnepp, M. Trunk, P. A. Walker, P. Winkler, and A. R. Maier, Lux – a laser-plasma driven undulator beamline, Nucl. Instr. Meth. Phys. Res. A **909**, 318 (2018).

[13] S. Jalas, I. Dornmair, R. Lehe, H. Vincenti, J. L. Vay, M. Kirchen, and A. R. Maier, Accurate modeling of plasma acceleration with arbitrary order pseudo-spectral particle-in-cell methods, Phys. Plasmas **24**, 033115 (2017).

[14] R. Brinkmann, N. Delbos, I. Dornmair, M. Kirchen, R. Assmann, C. Behrens, K. Floettmann, J. Grebenyuk, M. Gross, S. Jalas, T. Mehrling, A. Martinez de la Ossa, J. Osterhoff, B. Schmidt, V. Wacker, and A. R. Maier, Chirp Mitigation of Plasma-Accelerated Beams by a Modulated Plasma Density, Phys. Rev. Lett. **118**, 214801 (2017).

[15] M. Kirchen, R. Lehe, B. B. Godfrey, I. Dornmair, S. Jalas, K. Peters, J. L. Vay, and A. R. Maier, Stable discrete representation of relativistically drifting plasmas, Phys. Plasmas **23**, 100704 (2016).

[16] F. R. Elder, A. M. Gurewitsch, R. V. Langmuir, and H. C. Pollock, Radiation from electrons in a synchrotron, Phys. Rev. **71**, 829–830 (1947).

[17] E. Rowe and F. E. Mills, Tantalus. 1. a dedicated storage ring synchrotron radiation source, Part. Accel. **4**, 211–227 (1973).

[18] C. G. Schroer, H.-C. Wille, O. H. Seeck, K. Bagschik, H. Schulte-Schrepping, M. Tischer, H. Graafsma, W. Laasch, K. Baev, S. Klumpp, *et al.*, The synchrotron radiation source PETRA III and its future ultra-low-emittance upgrade PETRA IV, The European Physical Journal Plus **137**, 1312 (2022).

[19] A. Kondratenko and E. Saldin, Generating of coherent radiation by a relativistic electron beam in an ondulator, Part. Accel. **10**, 207–216 (1980).

[20] R. Bonifacio, C. Pellegrini, and L. Narducci, Collective instabilities and high-gain regime free electron laser, in *AIP conference proceedings*, Vol. 118 (American Institute of Physics, 1984) pp. 236–259.

[21] B. W. McNeil and N. R. Thompson, X-ray free-electron lasers, Nature photonics **4**, 814–821 (2010).

[22] J. Andruszkow, B. Aune, V. Ayvazyan, N. Baboi, R. Bakker, V. Balakin, D. Barni, A. Bazhan, M. Bernard, A. Bosotti, J. C. Bourdon, W. Brefeld, R. Brinkmann, *et al.*, First observation of self-amplified spontaneous emission in a free-electron laser at 109 nm wavelength, Phys. Rev. Lett. **85**, 3825–3829 (2000).

[23] P. Emma, R. Akre, J. Arthur, R. Bionta, C. Bostedt, J. Bozek, A. Brachmann, P. Bucksbaum, R. Coffee, F.-J. Decker, *et al.*, First lasing and operation of an ångstrom-wavelength free-electron laser, nature photonics **4**, 641–647 (2010).

[24] W. Decking, S. Abeghyan, P. Abramian, A. Abramsky, A. Aguirre, C. Albrecht, P. Alou, M. Altarelli, P. Altmann, K. Amyan, *et al.*, A mhz-repetition-rate hard x-ray free-electron laser driven by a superconducting linear accelerator, Nature photonics **14**, 391–397 (2020).

[25] C. Lamberti, The use of synchrotron radiation techniques in the characterization of strained semiconductor heterostructures and thin films, Surface Science Reports **53**, 1–197 (2004).

[26] B. Beckhoff, R. Fliegauf, M. Kolbe, M. Müller, J. Weser, and G. Ulm, Reference-free total reflection x-ray fluorescence analysis of semiconductor surfaces with synchrotron radiation, Analytical Chemistry **79**, 7873–7882 (2007).

[27] P. Hermann, A. Hoehl, G. Ulrich, C. Fleischmann, A. Hermelink, B. Kästner, P. Patoka, A. Hornemann, B. Beckhoff, E. Rühl, *et al.*, Characterization of semiconductor materials using synchrotron radiation-based near-field infrared microscopy and nano-ftir spectroscopy, Optics express **22**, 17948–17958 (2014).

[28] K. Westesen, B. Siekmann, and M. H. Koch, Investigations on the physical state of lipid nanoparticles by synchrotron radiation x-ray diffraction, International journal of pharmaceutics **93**, 189–199 (1993).

[29] T. F. Custódio, H. Das, D. J. Sheward, L. Hanke, S. Pazicky, J. Pieprzyk, M. Sorgenfrei, M. A. Schroer, A. Y. Gruzinov, C. M. Jeffries, *et al.*, Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize sars-cov-2, Nature communications **11**, 5588 (2020).

[30] S. Günther, P. Y. Reinke, Y. Fernández-García, J. Lieske, T. J. Lane, H. M. Ginn, F. H. Koua, C. Ehrt, W. Ewert, D. Oberthuer, *et al.*, X-ray screening identifies active site and allosteric inhibitors of sars-cov-2 main protease, Science **372**, 642–646 (2021).

[31] Z.-H. Loh, G. Doumy, C. Arnold, L. Kjellsson, S. Southworth, A. Al Haddad, Y. Kumagai, M.-F. Tu, P. Ho, A. March, *et al.*, Observation of the fastest chemical processes in the radiolysis of water, Science **367**, 179–182 (2020).

[32] K. S. Zinchenko, F. Ardana-Lamas, I. Seidu, S. P. Neville, J. van der Veen, V. U. Lanfaloni, M. S. Schuurman, and H. J. Wörner, Sub-7-femtosecond conical-intersection dynamics probed at the carbon k-edge, Science **371**, 489–494 (2021).

[33] H. N. Chapman, P. Fromme, A. Barty, T. A. White, R. A. Kirian, A. Aquila, M. S. Hunter, J. Schulz, D. P. DePonte, U. Weierstall, *et al.*, Femtosecond x-ray protein nanocrystallography, Nature **470**, 73–77 (2011).

[34] C. Kupitz, S. Basu, I. Grotjohann, R. Fromme, N. A. Zatsepin, K. N. Rendek, M. S. Hunter, R. L. Shoeman, T. A. White, D. Wang, *et al.*, Serial time-resolved crystallography of photosystem ii using a femtosecond x-ray laser, Nature **513**, 261–265 (2014).

[35] R. Boll, J. M. Schäfer, B. Richard, K. Fehre, G. Kastirke, Z. Jurek, M. S. Schöffler, M. M. Abdullah, N. Anders, T. M. Baumann, *et al.*, X-ray multiphoton-induced coulomb explosion images complex single molecules, Nature Physics **18**, 423–428 (2022).

[36] S. Frisia, A. Borsato, I. J. Fairchild, and J. Susini, Variations in atmospheric sulphate recorded in stalagmites by synchrotron micro-xrf and xanes analyses, Earth and Planetary Science Letters **235**, 729–740 (2005).

[37] R. Molinder, T. Comyn, N. Hondow, J. Parker, and V. Dupont, In situ x-ray diffraction of cao based co 2 sorbents, Energy & Environmental Science **5**, 8958–8969 (2012).

[38] L. M. Rossbach, D. A. Brede, G. Nuyts, S. Cagno, R. M. S. Olsson, D. H. Oughton, G. Falkenberg, K. Janssens, and O. C. Lind, Synchrotron xrf analysis identifies cerium accumulation colocalized with pharyngeal deformities in ceo2 np-exposed caenorhabditis elegans, Environmental Science & Technology **56**, 5081–5089 (2022).

[39] T. Tajima and J. M. Dawson, Laser electron accelerator, Phys. Rev. Lett. **43**, 267 (1979).

[40] E. Esarey, C. B. Schroeder, and W. P. Leemans, Physics of laser-driven plasma-based electron accelerators, Rev. Mod. Phys. **81**, 1229 (2009).

[41] D. Strickland and G. Mourou, Compression of amplified chirped optical pulses, Opt. Commun. **56**, 219 (1985).

[42] C. G. R. Geddes, C. Toth, J. van Tilborg, E. Esarey, C. B. Schroeder, D. Bruhwiler, C. Nieter, J. Cary, and W. P. Leemans, High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding, Nature **431**, 538 (2004).

[43] J. Faure, Y. Glinec, A. Pukhov, S. Kiselev, S. Gordienko, E. Lefebvre, J. P. Rousseau, F. Burgy, and V. Malka, A laser–plasma accelerator producing monoenergetic electron beams, Nature **431**, 541 (2004).

[44] S. P. D. Mangles, C. D. Murphy, Z. Najmudin, A. G. R. Thomas, J. L. Collier, A. E. Dangor, E. J. Divall, P. S. Foster, J. G. Gallacher, C. J. Hooker, D. A. Jaroszynski, A. J. Langley, W. B. Mori, P. A. Norreys, F. S. Tsung, *et al.*, Monoenergetic beams of relativistic electrons from intense laser–plasma interactions, Nature **431**, 535 (2004).

[45] A. Buck, M. Nicolai, K. Schmid, C. M. Sears, A. Sävert, J. M. Mikhailova, F. Krausz, M. C. Kaluza, and L. Veisz, Real-time observation of laser-driven electron acceleration, Nat. Phys. **7**, 543 (2011).

[46] O. Lundh, J. Lim, C. Rechatin, L. Ammoura, A. Ben-Ismaïl, X. Davoine, G. Gallot, J. P. Goddet, E. Lefebvre, V. Malka, and J. Faure, Few femtosecond, few kiloampere electron bunch produced by a laser-plasma accelerator, Nat. Phys. **7**, 219 (2011).

[47] J. P. Couperus, R. Pausch, A. Köhler, O. Zarini, J. M. Krämer, M. Garten, A. Huebl, R. Gebhardt, U. Helbig, S. Bock, K. Zeil, A. Debus, M. Bussmann, U. Schramm, and A. Irman, Demonstration of a beam loaded nanocoulomb-class laser wakefield accelerator, Nat. Commun. **8**, 487 (2017).

[48] R. Weingartner, S. Raith, A. Popp, S. Chou, J. Wenz, K. Khrennikov, M. Heigoldt, A. R. Maier, N. Kajumba, M. Fuchs, B. Zeitler, F. Krausz, S. Karsch, and F. Grüner, Ultralow emittance electron beams from a laser-wakefield accelerator, Phys. Rev. ST Accel. Beams **15**, 111302 (2012).

[49] G. R. Plateau, C. G. R. Geddes, D. B. Thorn, M. Chen, C. Benedetti, E. Esarey, A. J. Gonsalves, N. H. Matlis, K. Nakamura, C. B. Schroeder, S. Shiraishi, T. Sokollik, J. van Tilborg, C. Toth, S. Trotsenko, *et al.*, Low-Emittance Electron Bunches from a Laser-Plasma Accelerator Measured using Single-Shot X-Ray Spectroscopy, Phys. Rev. Lett. **109**, 064802 (2012).

[50] S. K. Barber, J. van Tilborg, C. B. Schroeder, R. Lehe, H.-E. Tsai, K. K. Swanson, S. Steinke, K. Nakamura, C. G. R. Geddes, C. Benedetti, E. Esarey, and W. P. Leemans, Measured emittance dependence on the injection method in laser plasma accelerators, Phys. Rev. Lett. **119**, 104801 (2017).

[51] C. Rechatin, J. Faure, A. Ben-Ismail, J. Lim, R. Fitour, A. Specka, H. Videau, A. Tafzi, F. Burgy, and V. Malka, Controlling the Phase-Space Volume of Injected Electrons in a Laser-Plasma Accelerator, Phys. Rev. Lett. **102**, 164801 (2009).

[52] L. T. Ke, K. Feng, W. T. Wang, Z. Y. Qin, C. H. Yu, Y. Wu, Y. Chen, R. Qi, Z. J. Zhang, Y. Xu, X. J. Yang, Y. X. Leng, J. S. Liu, R. X. Li, and Z. Z. Xu, Near-gev electron beams at a few per-mille level from a laser wakefield accelerator via density-tailored plasma, Phys. Rev. Lett. **126**, 214801 (2021).

[53] W. P. Leemans, B. Nagler, A. J. Gonsalves, C. Tóth, K. Nakamura, C. G. Geddes, E. Esarey, C. B. Schroeder, and S. M. Hooker, Gev electron beams from a centimetre-scale accelerator, Nat. Phys. **2**, 696 (2006).

[54] X. Wang, R. Zgadzaj, N. Fazel, Z. Li, S. A. Yi, X. Zhang, W. Henderson, Y. Y. Chang, R. Korzekwa, H. E. Tsai, C. H. Pai, H. Quevedo, G. Dyer, E. Gaul, M. Martinez, *et al.*, Quasi-monoenergetic laser-plasma acceleration of electrons to 2 gev, Nature Communications **4**, 1988 (2013).

[55] W. P. Leemans, A. J. Gonsalves, H.-S. Mao, K. Nakamura, C. Benedetti, C. B. Schroeder, C. Tóth, J. Daniels, D. E. Mittelberger, S. S. Bulanov, J.-L. Vay, C. G. R. Geddes, and E. Esarey, Multi-GeV Electron Beams from Capillary-Discharge-Guided Subpetawatt Laser Pulses in the Self-Trapping Regime, Phys. Rev. Lett. **113**, 245002 (2014).

[56] A. J. Gonsalves, K. Nakamura, J. Daniels, C. Benedetti, C. Pieronek, T. C. H. de Raadt, S. Steinke, J. H. Bin, S. S. Bulanov, J. van Tilborg, C. G. R. Geddes, C. B. Schroeder, C. Tóth, E. Esarey, K. Swanson, *et al.*, Petawatt Laser Guiding and Electron Beam Acceleration to 8 GeV in a Laser-Heated Capillary Discharge Waveguide, Phys. Rev. Lett. **122**, 084801 (2019).

[57] W. Wang, K. Feng, L. Ke, C. Yu, Y. Xu, R. Qi, Y. Chen, Z. Qin, Z. Zhang, M. Fang, *et al.*, Free-electron lasing at 27 nanometres based on a laser wakefield accelerator, Nature **595**, 516–520 (2021).

[58] P. Gibbon, *Short Pulse Laser Interactions with Matter* (World Scientific, 2005).

[59] P. Gibbon, Introduction to plasma physics, CERN Yellow Reports 10.5170/CERN-2016-001.51 (2016).

[60] F. Chen, *Introduction to Plasma Physics and Controlled Fusion* (Springer, 1984).

[61] J. Jackson and P. Jackson, *Classical Electrodynamics* (Wiley, 1998).

[62] P. Sprangle, E. Esarey, and A. Ting, Nonlinear interaction of intense laser pulses in plasmas, Phys. Rev. A **41**, 4463 (1990).

[63] V. Berezhiani and I. Murusidze, Interaction of highly relativistic short laser pulses with plasmas and nonlinear wake-field generation, Physica Scripta **45**, 87 (1992).

[64] J. M. Dawson, Nonlinear Electron Oscillations in a Cold Plasma, Phys. Rev. **113**, 383 (1959).

[65] J. M. Dawson, Particle simulation of plasmas, Rev. Mod. Phys. **55**, 403 (1983).

[66] W. Lu, C. Huang, M. Zhou, W. B. Mori, and T. Katsouleas, Nonlinear theory for relativistic plasma wakefields in the blowout regime, Phys. Rev. Lett. **96**, 165002 (2006).

[67] E. Esarey and M. Pilloff, Trapping and acceleration in nonlinear plasma waves, Physics of Plasmas **2**, 1432 (1995).

[68] T. Esirkepov, S. V. Bulanov, M. Yamagiwa, and T. Tajima, Electron, Positron, and Photon Wakefield Acceleration: Trapping, Wake Overtaking, and Ponderomotive Acceleration, Phys. Rev. Lett. **96**, 014803 (2006).

[69] M. Kirchen, *Novel Concepts for the Simulation and Experimental Demonstration of High-Quality Laser-Plasma Acceleration*, Ph.D. thesis, Universität Hamburg (2021).

[70] S. Steinke, J. Van Tilborg, C. Benedetti, C. Geddes, C. Schroeder, J. Daniels, K. Swanson, A. Gonsalves, K. Nakamura, N. Matlis, *et al.*, Multistage coupling of independent laser-plasma accelerators, Nature **530**, 190–193 (2016).

[71] S. V. Bulanov, F. Pegoraro, A. M. Pukhov, and A. S. Sakharov, Transverse-wake wave breaking, Phys. Rev. Lett. **78**, 4205 (1997).

[72] S. Y. Kalmykov, L. M. Gorbunov, P. Mora, and G. Shvets, Injection, trapping, and acceleration of electrons in a three-dimensional nonlinear laser wakefield, Physics of plasmas **13** (2006).

[73] C. B. Schroeder and E. Esarey, Relativistic warm plasma theory of nonlinear laser-driven electron plasma waves, Phys. Rev. E **81**, 056403 (2010).

[74] S. Kalmykov, S. A. Yi, V. Khudik, and G. Shvets, Electron self-injection and trapping into an evolving plasma bubble, Phys. Rev. Lett. **103**, 135004 (2009).

[75] S. Bulanov, N. Naumova, F. Pegoraro, and J. Sakai, Particle injection into the wave acceleration phase due to nonlinear wake wave breaking, Phys. Rev. E **58**, R5257 (1998).

[76] H. Suk, N. Barov, J. B. Rosenzweig, and E. Esarey, Plasma Electron Trapping and Acceleration in a Plasma Wake Field Using a Density Transition, Phys. Rev. Lett. **86**, 1011 (2001).

[77] M. V. Ammosov, N. B. Delone, and V. P. Krainov, Tunnel ionization of complex atoms and atomic ions in an electromagnetic field, Sov. Phys. JETP **64**, 1191 (1986).

[78] M. Chen, E. Cormier-Michel, C. Geddes, D. Bruhwiler, L. Yu, E. Esarey, C. Schroeder, and W. Leemans, Numerical modeling of laser tunneling ionization in explicit particle-in-cell codes, J. Comput. Phys **236**, 220 (2013).

[79] M. Chen, E. Esarey, C. B. Schroeder, C. G. R. Geddes, and W. P. Leemans, Theory of ionization-induced trapping in laser-plasma accelerators, Phys. Plasmas **19**, 033101 (2012).

[80] C. B. Schroeder, J.-L. Vay, E. Esarey, S. S. Bulanov, C. Benedetti, L.-L. Yu, M. Chen, C. G. R. Geddes, and W. P. Leemans, Thermal emittance from ionization-induced trapping in plasma accelerators, Phys. Rev. ST Accel. Beams **17**, 101301 (2014).

[81] E. Esarey, R. F. Hubbard, W. P. Leemans, A. Ting, and P. Sprangle, Electron injection into plasma wakefields by colliding laser pulses, Phys. Rev. Lett. **79**, 2682 (1997).

[82] L.-L. Yu, E. Esarey, C. B. Schroeder, J.-L. Vay, C. Benedetti, C. G. R. Geddes, M. Chen, and W. P. Leemans, Two-color laser-ionization injection, Phys. Rev. Lett. **112**, 125001 (2014).

[83] P. Sprangle, E. Esarey, J. Krall, and G. Joyce, Propagation and guiding of intense laser pulses in plasmas, Phys. Rev. Lett. **69**, 2200 (1992).

[84] E. Esarey, P. Sprangle, J. Krall, and A. Ting, Self-focusing and guiding of short laser pulses in ionizing gases and plasmas, IEEE Journal of Quantum Electronics **33**, 1879 (1997).

[85] P. Sprangle, E. Esarey, A. Ting, and G. Joyce, Laser wakefield acceleration and relativistic optical guiding, Applied Physics Letters **53**, 2146 (1988).

[86] E. Esarey, B. A. Shadwick, C. B. Schroeder, and W. P. Leemans, Nonlinear Pump Depletion and Electron Dephasing in Laser Wakefield Accelerators, AIP Conference Proceedings **737**, 578 (2004).

[87] W. Rittershofer, C. Schroeder, E. Esarey, F. Grüner, and W. Leemans, Tapered plasma channels to phase-lock accelerating and focusing forces in laser-plasma accelerators, Physics of Plasmas **17** (2010).

[88] B. A. Shadwick, C. B. Schroeder, and E. Esarey, Nonlinear laser energy depletion in laser-plasma accelerators, Physics of Plasmas **16**, 056704 (2009).

[89] S. Van der Meer, *Improving the power efficiency of the plasma wakefield accelerator*, Tech. Rep. Ps-85-65 (Cern, 1985).

[90] T. C. Katsouleas, S. Wilks, P. Chen, J. M. Dawson, and J. J. Su, Beam loading in plasma accelerators, Part. Accel. **22**, 81 (1987).

[91] M. Tzoufras, W. Lu, F. S. Tsung, C. Huang, W. B. Mori, T. Katsouleas, J. Vieira, R. A. Fonseca, and L. O. Silva, Beam loading in the nonlinear regime of plasma-based acceleration, Phys. Rev. Lett. **101**, 145002 (2008).

[92] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis* (Chapman and Hall/CRC, 1995).

[93] J. Kruschke, *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (Academic Press, 2014).

[94] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning.* (MIT Press, 2006).

[95] B. Minasny and A. B. McBratney, The Matérn function as a general model for soil variograms, Geoderma **128**, 192–207 (2005), pedometrics 2003.

[96] D. R. Jones, M. Schonlau, and W. J. Welch, Efficient Global Optimization of Expensive Black-Box Functions, Journal of Global Optimization **13**, 455–492 (1998).

[97] P. I. Frazier, A Tutorial on Bayesian Optimization, arXiv e-prints , arXiv:1807.02811 (2018), arXiv:1807.02811 [stat.ML] .

[98] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, A Limited Memory Algorithm for Bound Constrained Optimization, SIAM Journal on Scientific Computing **16**, 1190–1208 (1995), https://doi.org/10.1137/0916069 .

[99] A. D. Bull, Convergence rates of efficient global optimization algorithms., Journal of Machine Learning Research **12** (2011).

[100] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design, in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10 (Omnipress, Madison, WI, USA, 2010) p. 1015–1022.

[101] J. Duris, D. Kennedy, A. Hanuka, J. Shtalenkova, A. Edelen, P. Baxevanis, A. Egger, T. Cope, M. McIntire, S. Ermon, and D. Ratner, Bayesian optimization of a free-electron laser, Phys. Rev. Lett. **124**, 124801 (2020).

[102] P. Frazier, W. Powell, and S. Dayanik, The knowledge-gradient policy for correlated normal beliefs, INFORMS journal on Computing **21**, 599–613 (2009).

[103] M. Balandat, B. Karrer, D. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization, in *Advances in Neural Information Processing Systems*, Vol. 33, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc., 2020) pp. 21524–21538.

[104] J. Villemonteix, E. Vazquez, M. Sidorkiewicz, and E. Walter, Global optimization of expensive-to-evaluate functions: an empirical comparison of two sampling criteria, Journal of Global Optimization **43**, 373–389 (2009).

[105] B. Letham, B. Karrer, G. Ottoni, and E. Bakshy, Constrained Bayesian optimization with noisy experiments, Bayesian Analysis **14**, 495–519 (2019).

[106] M. Emmerich, K. Giannakoglou, and B. Naujoks, Single- and multiobjective evolutionary optimization assisted by Gaussian random field metamodels, IEEE Transactions on Evolutionary Computation **10**, 421–439 (2006).

[107] I. Hupkens, A. Deutz, K. Yang, and M. Emmerich, Faster exact algorithms for computing expected hypervolume improvement, in *international conference on evolutionary multi-criterion optimization* (Springer, 2015) pp. 65–79.

[108] M. Emmerich, K. Yang, A. Deutz, H. Wang, and C. M. Fonseca, A multicriteria generalization of bayesian global optimization, in *Advances in stochastic and deterministic global optimization* (Springer, 2016) pp. 229–242.

[109] K. Yang, M. Emmerich, A. Deutz, and C. M. Fonseca, Computing 3-D Expected Hypervolume Improvement and Related Integrals in Asymptotically Optimal Time, in *Evolutionary Multi-Criterion Optimization*, edited by H. Trautmann, G. Rudolph, K. Klamroth, O. Schütze, M. Wiecek, Y. Jin, and C. Grimme (Springer International Publishing, Cham, 2017) pp. 685–700.

[110] S. Daulton, M. Balandat, and E. Bakshy, Differentiable Expected Hypervolume Improvement for Parallel Multi-Objective Bayesian Optimization, in *Advances in Neural Information Processing Systems*, Vol. 33, edited by H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin (Curran Associates, Inc., 2020) pp. 9851–9864.

[111] S. Daulton, M. Balandat, and E. Bakshy, Parallel Bayesian Optimization of Multiple Noisy Objectives with Expected Hypervolume Improvement, Advances in Neural Information Processing Systems **34** (2021).

[112] J. Knowles, ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems, IEEE Transactions on Evolutionary Computation **10**, 50–66 (2006).

[113] D. Hernández-Lobato, J. Hernandez-Lobato, A. Shah, and R. Adams, Predictive entropy search for multi-objective bayesian optimization, in *International conference on machine learning* (PMLR, 2016) pp. 1492–1501.

[114] A. J. Gonsalves, T. P. Rowlands-Rees, B. H. P. Broks, J. J. A. M. van der Mullen, and S. M. Hooker, Transverse Interferometry of a Hydrogen-Filled Capillary Discharge Waveguide, Phys. Rev. Lett. **98**, 025002 (2007).

[115] R. J. Shalloo, C. Arran, L. Corner, J. Holloway, J. Jonnerby, R. Walczak, H. M. Milchberg, and S. M. Hooker, Hydrodynamic optical-field-ionized plasma channels, Phys. Rev. E **97**, 053203 (2018).

[116] A. Buck, J. Wenz, J. Xu, K. Khrennikov, K. Schmid, M. Heigoldt, J. M. Mikhailova, M. Geissler, B. Shen, F. Krausz, S. Karsch, and L. Veisz, Shock-front injector for high-quality laser-plasma acceleration, Phys. Rev. Lett. **110**, 185006 (2013).

[117] K. K. Swanson, H.-E. Tsai, S. K. Barber, R. Lehe, H.-S. Mao, S. Steinke, J. van Tilborg, K. Nakamura, C. G. R. Geddes, C. B. Schroeder, E. Esarey, and W. P. Leemans, Control of tunable, monoenergetic laser-plasma-accelerated electron beams using a shock-induced density downramp injector, Phys. Rev. Accel. Beams **20**, 051301 (2017).

[118] W. T. Wang, W. T. Li, J. S. Liu, Z. J. Zhang, R. Qi, C. H. Yu, J. Q. Liu, M. Fang, Z. Y. Qin, C. Wang, Y. Xu, F. X. Wu, Y. X. Leng, R. X. Li, and Z. Z. Xu, High-Brightness High-Energy Electron Beams from a Laser Wakefield Accelerator via Energy Chirp Control, Phys. Rev. Lett. **117**, 124801 (2016).

[119] S. Bohlen, J. C. Wood, T. Brümmer, F. Grüner, C. A. Lindstrøm, M. Meisel, T. Staufer, R. D'Arcy, K. Põder, and J. Osterhoff, Stability of ionization-injection-based laser-plasma accelerators, Phys. Rev. Accel. Beams **25**, 031301 (2022).

[120] S. Kuschel, M. B. Schwab, M. Yeung, D. Hollatz, A. Seidel, W. Ziegler, A. Sävert, M. C. Kaluza, and M. Zepf, Controlling the self-injection threshold in laser wakefield accelerators, Phys. Rev. Lett. **121**, 154801 (2018).

[121] M. Mirzaie, S. Li, M. Zeng, N. A. M. Hafz, M. Chen, G. Y. Li, Q. J. Zhu, H. Liao, T. Sokollik, F. Liu, Y. Y. Ma, L. M. Chen, Z. M. Sheng, and J. Zhang, Demonstration of self-truncated ionization injection for GeV electron beams, Sci. Rep. **5**, 14659 (2015).

[122] B. B. Pollock, C. E. Clayton, J. E. Ralph, F. Albert, A. Davidson, L. Divol, C. Filip, S. H. Glenzer, K. Herpoldt, W. Lu, K. A. Marsh, J. Meinecke, W. B. Mori, A. Pak, T. C. Rensink, *et al.*, Demonstration of a Narrow Energy Spread, ~ 0.5 GeV Electron Beam from a Two-Stage Laser Wakefield Accelerator, Phys. Rev. Lett. **107**, 045001 (2011).

[123] A. J. Gonsalves, K. Nakamura, C. Benedetti, C. V. Pieronek, S. Steinke, J. H. Bin, S. S. Bulanov, J. van Tilborg, C. G. R. Geddes, C. B. Schroeder, J. Daniels, C. Tóth, L. Obst-Huebl, R. G. W. van den Berg, G. Bagdasarov, *et al.*, Laser-heated capillary discharge plasma waveguides for electron acceleration to 8 GeV, Physics of Plasmas **27**, 053102 (2020).

[124] R. Lehe, M. Kirchen, I. A. Andriyash, B. B. Godfrey, and J.-L. Vay, A spectral, quasi-cylindrical and dispersion-free Particle-In-Cell algorithm, Comput. Phys. Commun. **203**, 66 (2016).

[125] A. Lifschitz, X. Davoine, E. Lefebvre, J. Faure, C. Rechatin, and V. Malka, Particle-in-Cell modelling of laser–plasma interaction using Fourier decomposition, J. Comput. Phys **228**, 1803 (2009).

[126] J. L. Vay, Noninvariance of space- and time-scale ranges under a Lorentz transformation and the implications for the study of relativistic interactions, Phys. Rev. Lett. **98**, 130405 (2007).

[127] R. Lehe, M. Kirchen, B. B. Godfrey, A. R. Maier, and J. L. Vay, Elimination of numerical Cherenkov instability in flowing-plasma particle-in-cell simulations by using Galilean coordinates, Phys. Rev. E **94**, 053305 (2016).

[128] M. Santarsiero, D. Aiello, R. Borghi, and S. Vicalvi, Focusing of axially symmetric flattened Gaussian beams, Journal of Modern Optics **44**, 633 (1997).

[129] R. Shalloo, S. Dann, J.-N. Gruse, C. Underwood, A. Antoine, C. Arran, M. Backhouse, C. Baird, M. Balcazar, N. Bourgeois, *et al.*, Automation and control of laser wakefield accelerators using Bayesian optimization, Nature communications **11**, 1–8 (2020).

[130] V. Leroux, *On the wavefront of ultrahigh intensity lasers: spatial contrast and gratings deformation*, Ph.D. thesis, Universität Hamburg (2018).

[131] C. Werle, *First Undulator Experiments at the LUX Beamline: Towards a Plasma-Based Accelerator as a Soft X-ray Source*, Ph.D. thesis, Universität Hamburg (2019).

[132] P. Winkler, *Emittance Measurements at Laser-Wakefield Accelerators*, Ph.D. thesis, Universität Hamburg (2019).

[133] P. Messner, *Tunable Laser-Plasma Acceleration with Ionization Injection*, Ph.D. thesis, Universität Hamburg (2020).

[134] K. F. Wall and A. Sanchez, Titanium sapphire lasers, The Lincoln laboratory journal **3**, 447 (1990).

[135] V. Leroux, T. Eichner, and A. R. Maier, Description of spatio-temporal couplings from heat-induced compressor grating deformation, Opt. Express **28**, 8257–8265 (2020).

[136] A. Jullien, O. Albert, F. Burgy, G. Hamoniaux, J.-P. Rousseau, J.-P. Chambaret, F. Augé-Rochereau, G. Chériaux, J. Etchepare, N. Minkovski, and S. M. Saltiel, 10-10 temporal contrast for femtosecond ultraintense lasers by cross-polarized wave generation, Opt. Lett. **30**, 920 (2005).

[137] A. Dubietis, G. Jonušauskas, and A. Piskarskas, Powerful femtosecond pulse generation by chirped and stretched pulse parametric amplification in bbo crystal, Optics Communications **88**, 437–440 (1992).

[138] I. Ross, P. Matousek, M. Towrie, A. Langley, and J. Collier, The prospects for ultrashort pulse duration and ultrahigh intensity using optical parametric chirped pulse amplifiers, Optics Communications **144**, 125–133 (1997).

[139] T. Eichner, T. Hülsenbusch, J. Dirkwinkel, T. Lang, L. Winkelmann, G. Palmer, and A. R. Maier, Spatio-spectral couplings in saturated collinear opcpa, Opt. Express **30**, 3404–3415 (2022).

[140] P. Tournois, Acousto-optic programmable dispersive filter for adaptive compensation of group delay time dispersion in laser systems, Opt. Commun. **140**, 245 (1997).

[141] R. V. Shack, Production and use of a lecticular hartmann screen, in *Program of the 1971 Spring Meeting of the Optical Society of America*, Vol. 61 (J. Opt. Soc. Am., 1971) p. 656.

[142] T. Oksenhendler, S. Coudreau, N. Forget, V. Crozatier, S. Grabielle, R. Herzog, O. Gobert, and D. Kaplan, Self-referenced spectral interferometry, Appl. Phys. B **99**, 7 (2010).

[143] D. Lipka *et al.*, Very first experience with the standard diagnostics at the european xfel, in *Proc. of 8th IPAC* (Copenhagen, Denmark, 2017) p. 180.

[144] K. P. Blum, *Calibration of an Electro-Dipole Electron Spectrometer for the LUX Laser-Plasma-Accelerator Beamline Upgrade*, Bachelor's thesis, Universität Hamburg (2020).

[145] D. Lipka, J. Lund-Nielsen, and M. Seebach, Resonator for Charge Measurement at REGAE (2nd International Beam Instrumentation Conference, Oxford (United Kingdom), 16 Sep 2013 - 19 Sep 2013, 2013).

[146] L. Hübner and P. Winkler, lux tracker (2023).

[147] K. Floettmann, Some basic features of the beam emittance, Phys. Rev. ST Accel. Beams **6**, 034202 (2003).

[148] T. Mehrling, J. Grebenyuk, F. S. Tsung, K. Floettmann, and J. Osterhoff, Transverse emittance growth in staged laser-wakefield acceleration, Phys. Rev. ST Accel. Beams **15**, 111303 (2012).

[149] O. Hensler and K. Rehlich, Doocs: A distributed object oriented control system, in *Proceedings of XV Workshop on Charged Particle Accelerators, Protvino* (1996).

[150] P. Winkler, S. Jalas, M. Kirchen, and L. Hübner, lux pipeline (2023).

[151] S. Corde, K. Ta Phuoc, G. Lambert, R. Fitour, V. Malka, A. Rousse, A. Beck, and E. Lefebvre, Femtosecond x rays from laser-plasma accelerators, Rev. Mod. Phys. **85**, 1 (2013).

[152] C. B. Schroeder, E. Esarey, C. G. R. Geddes, C. Benedetti, and W. P. Leemans, Physics considerations for laser-plasma linear colliders, Phys. Rev. ST Accel. Beams **13**, 101301 (2010).

[153] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization* (John Wiley & Sons, 2015).

[154] A. J. Gonsalves, K. Nakamura, C. Lin, D. Panasenko, S. Shiraishi, T. Sokollik, C. Benedetti, C. B. Schroeder, C. G. R. Geddes, J. van Tilborg, J. Osterhoff, E. Esarey, C. Toth, and W. P. Leemans, Tunable laser plasma accelerator based on longitudinal density tailoring, Nat. Phys. **7**, 862 (2011).

[155] J. Götzfried, A. Döpp, M. F. Gilljohann, F. M. Foerster, H. Ding, S. Schindler, G. Schilling, A. Buck, L. Veisz, and S. Karsch, Physics of high-charge electron beams in laser-plasma wakefields, Phys. Rev. X **10**, 041015 (2020).

[156] K. P. Murphy, *Machine learning: a probabilistic perspective* (MIT press, 2012) pp. 492–493.

[157] H. Ye, Y. Gu, X. Zhang, S. Wang, F. Tan, J. Zhang, Y. Yang, Y. Yan, Y. Wu, W. Huang, *et al.*, Fast optimization for betatron radiation from laser wakefield acceleration based on bayesian optimization, Results in Physics **43**, 106116 (2022).

[158] F. Irshad, S. Karsch, and A. Döpp, Multi-objective and multi-fidelity bayesian optimization of laser-plasma acceleration, Phys. Rev. Res. **5**, 013063 (2023).

[159] S. Bohlen, Z. Gong, M. J. Quin, M. Tamburini, and K. Põder, Colliding pulse injection of polarized electron bunches in a laser-plasma accelerator, arXiv preprint arXiv:2304.02922 (2023).

[160] B. Loughran, M. Streeter, H. Ahmed, S. Astbury, M. Balcazar, M. Borghesi, N. Bourgeois, C. Curry, S. Dann, S. DiIorio, *et al.*, Automated control and optimization of laser-driven ion acceleration, High Power Laser Science and Engineering **11**, e35 (2023).

[161] J. Goodman, M. King, E. Dolier, R. Wilson, R. Gray, and P. McKenna, Optimization and control of synchrotron emission in ultraintense laser–solid interactions using machine learning, High Power Laser Science and Engineering **11**, e34 (2023).

[162] S. A. Antipov, I. V. Agapov, R. Brinkmann, A. F. Pousa, M. A. Jebramcik, A. M. de la Ossa, and M. Thévenet, First start-to-end simulations of the 6 gev laser-plasma injector at desy, in *Proc. IPAC'22*, International Particle Accelerator Conference No. 13 (JACoW Publishing, Geneva, Switzerland, 2022) pp. 1757–1760.

[163] C. Adolphsen, D. Angal-Kalinin, T. Arndt, M. Arnold, R. Assmann, B. Auchmann, K. Aulenbacher, A. Ballarino, B. Baudouy, P. Baudrenghien, *et al.*, European strategy for particle physics–accelerator r&d roadmap, arXiv preprint arXiv:2201.07895 10.48550/arXiv.2201.07895 (2022).

[164] C. Schroeder, F. Albert, C. Benedetti, J. Bromage, D. Bruhwiler, S. Bulanov, E. Campbell, N. Cook, B. Cros, M. Downer, *et al.*, Linear colliders based on laser-plasma accelerators, Journal of Instrumentation **18** (06), T06001.

[165] M. Fuchs, B. Shadwick, N. Vafaei-Najafabadi, A. Thomas, G. Andonian, M. Büscher, A. Lehrach, O. Apsimon, G. Xia, D. Filippetto, *et al.*, Snowmass whitepaper af6: Plasma-based particle sources, arXiv preprint arXiv:2203.08379 (2022).

[166] S. Biedron, L. Brouwer, D. Bruhwiler, N. Cook, A. Edelen, D. Filippetto, C.-K. Huang, A. Huebl, T. Katsouleas, N. Kuklev, *et al.*, Snowmass21 accelerator modeling community white paper, arXiv preprint arXiv:2203.08335 (2022).

[167] S. J. Russell and P. Norvig, *Artificial intelligence: a modern approach* (Malaysia; Pearson Education Limited,, 2016).

[168] J. Kaiser, C. Xu, A. Eichler, A. S. Garcia, O. Stein, E. Bründermann, W. Kuropka, H. Dinter, F. Mayet, T. Vinatier, *et al.*, Learning to do or learning while doing: Reinforcement learning and bayesian optimisation for online continuous tuning, arXiv preprint arXiv:2306.03739 (2023).

[169] L. Bottou, Online algorithms and stochastic approximations, in *Online Learning and Neural Networks*, edited by D. Saad (Cambridge University Press, Cambridge, UK, 1998) revised, oct 2012.

[170] A. Scheinker, A. Edelen, D. Bohler, C. Emma, and A. Lutman, Demonstration of model-independent control of the longitudinal phase space of electron beams in the linac-coherent light source with femtosecond resolution, Phys. Rev. Lett. **121**, 044801 (2018).

[171] D. Ratner, F. Christie, J. P. Cryan, A. Edelen, A. Lutman, and X. Zhang, Recovering the phase and amplitude of x-ray fel pulses using neural networks and differentiable models, Opt. Express **29**, 20336–20352 (2021).

[172] R. Roussel, J. P. Gonzalez-Aguilera, Y.-K. Kim, E. Wisniewski, W. Liu, P. Piot, J. Power, A. Hanuka, and A. Edelen, Turn-key constrained parameter space exploration for particle accelerators using bayesian active learning, Nature communications **12**, 5612 (2021).

[173] R. Roussel, A. Edelen, D. Ratner, K. Dubey, J. P. Gonzalez-Aguilera, Y. K. Kim, and N. Kuklev, Differentiable preisach modeling for characterization and optimization of particle accelerator systems with hysteresis, Phys. Rev. Lett. **128**, 204801 (2022).

[174] R. Roussel, A. Edelen, C. Mayes, D. Ratner, J. P. Gonzalez-Aguilera, S. Kim, E. Wisniewski, and J. Power, Phase space reconstruction from accelerator beam measurements using neural networks and differentiable simulations, Phys. Rev. Lett. **130**, 145001 (2023).

Figures and tables

List of Figures

List of Tables